The University of Texas Aerospace Crew Research Project

**Aircrews' Evaluations of Flight Deck Automation Training and Use: Measuring and Ameliorating Threats to Safety**[1]

Paul J. Sherman, PhD

Technical Report 97-2[2]
July 31, 1997

**ABSTRACT**

The present study examined 1,718 commercial airline pilots' evaluations of the training they received for use of aircraft automation, automated systems on their current aircraft, and their attitudes toward the use and management of automation. Examination of training ratings showed that, overall, roughly one-quarter of pilots felt that initial training did not adequately prepare them for operating their aircraft. Substantial differences in ratings of training efficacy were found across airlines, aircraft types, experience level, and exposure to discretionary opportunities for practice during training. Examination of automated equipment evaluations revealed that ratings of automation usability are related to ratings of training efficacy, implying that any evaluations of automated equipment must take training efficacy into account. Analyses also demonstrated differences across aircraft types on automation usability, quality of troubleshooting and problem solving, and awareness of aircraft energy state; some of these differences seem to be related to differences across aircraft manufacturers and some to differences in automation generation. Finally, analyses of pilots' attitudes toward management of automation showed relationships between the scales and measures of experience, perceptions of company policies regarding automation use, and a measure of respondents' need to avoid uncertain, ambiguous situations. Overall, these results allow identification of some potential threats to safety that reside in the crew-automation interface. They also suggest that crew-automation interaction can be conceptualized from the systems viewpoint – i.e., that crew-automation interaction is determined by multiple factors, including training quality, the automated equipment itself, and the organization's policies and procedures regarding automation use.

---

[1] This report is based on Paul Sheman's doctoral dissertation study of the same title.

[2] The present research was supported by FAA Grant 92-G-017, Robert L. Helmreich, Principal Investigator.

# INTRODUCTION

The successful management of complex, team-oriented processes such as aviation, industrial, maritime, and surgical endeavors requires that team members utilize a combination of cognitive, psychomotor, and interpersonal skills, often in situations of considerable risk. Automation, defined as the replacement of a human function, either manual or cognitive, with a machine function (Wiener, Chidester, Kanki, Palmer, Curry & Gregorich, 1991), has been deployed in these domains to prevent human error, aid team members in accomplishment of tasks, and to increase efficiency (Billings, 1997; Drury, 1996; Lee & Sandquist, 1996; Meshkati, 1996). Since the widespread application of automation in aviation beginning in the early 1980's, the overall accident rate has decreased slightly. However, this overall decrease masks the existence of a trend in incidents and accidents that are attributable in whole or in part to pilots' interaction with automated systems (Billings, 1997). The following are two examples of these types of accidents:

1. In December 1995, the crew of an American Airlines Boeing 757 descending through a mountain valley toward Cali, Columbia, attempted to route the aircraft toward their destination by entering into the flight management system (FMS) a substring of the code for a Cali navigational beacon. The computer's stored database of navigational beacons contained two very similar codes. One code denoted the beacon near Cali, which was several dozen miles ahead of the airplane. The other code corresponded to a beacon at the Bogota airport, several dozen miles *behind* the airplane. Presented by the FMS with a list of nearby beacons matching the inputted substring, the crew initiated an overlearned behavior, and selected the computer's first presented alternative – unfortunately, the FMS had presented the Bogota beacon first. The flight management computer dutifully began to turn the aircraft toward Bogota. Shortly after this, the aircraft crashed into the side of a mountain, killing all on board (Federal Aviation Administration [FAA], 1996).

2. A China Air A300-600 approaching Nagoya, Japan in April, 1994 crashed after the pilots spent several minutes struggling to land while the autopilot was attempting to ascend the aircraft and carry out a "go around" (i.e., an aborted landing). During the last minutes of flight the crew had the ability to completely disengage the autopilot, but did not exercise this option (Sherman & Wiener, 1995).

Unfortunately, there are a number of other examples of human-automation mishaps in aviation and other industries. In several instances, automated systems have behaved in accordance with their design specifications, but at inappropriate times (FAA, 1996). In other cases, crews have failed to detect or the automation has failed to inform crews of system malfunctions (Billings, 1997). Despite the best intentions of aircraft and software designers, as Wiener (1993b) observes, automation cannot completely eliminate human error, and may in some cases exacerbate it.

The amount of air traffic worldwide has more than doubled in the past twenty years, and this rate of growth is expected to continue through the first quarter of the 21st century (FAA, 1996). Although the accident rate has fallen steadily during this time, for the past several years it has remained roughly constant at 0.3 accidents per million flights (Sears, 1989). With the projected increase in air traffic, in the coming years there will be an unacceptable increase in the number of accidents, with a commercial jet accident occurring on average once every week or two (Weener, 1991). The only way to prevent this is to reduce the accident rate. The Air Transport Association of America ([ATA]; in press; 1989), the Air Line Pilots Association ([ALPA], 1996) and the FAA Human Factors Team (1996), a taskforce devoted to the study of flight deck automation and team performance, view better understanding of human-automation interaction as a key to reducing the accident rate, and have issued guidelines and identified areas of concern regarding automation design and use. It is hoped that these efforts will eventually influence the design of automation for future aircraft.

However, to better understand human-automation interaction in today's automated aircraft, it is particularly important to identify areas of concern regarding aircrews' present automation use. To accomplish this, more information is needed regarding how air crews view their training for automation, how they view the automation itself, and how they report using it in actual line operations. This knowledge is critical to reducing the accident rate through the design and implementation of better training philosophies and practices. Only when these areas of concern are identified and described can effective, empirically informed countermeasures be designed and implemented. The present study is an effort to aid in the development of such countermeasures by identifying and describing specific areas of concern regarding aircrews' evaluations of their own and the automation's performance, their evaluation of the training they received, and their attitudes toward automation use.

There are a number of hypotheses concerning what factors affect aircrews' use and evaluation of automation. Researchers have attributed differences in automation views and usage to cognitive factors (Amalberti, in press), experience level, pilots' prior experience with computers (Wiener, Chidester, Kanki, Palmer, Curry, & Gregorich, 1991), differences in aircraft and systems design (ALPA, 1996; Folkerts & Jorna, 1994), training design, organizational culture, and national culture (Helmreich & Sherman, 1994; Sherman & Helmreich, 1995; Sherman, Helmreich, & Merritt, in press). While this study will address some of these hypotheses, a primary goal of this work will be to describe as fully as possible air crew-automation interaction across a broad array of aircraft types, organizations, and air crew demographics. Thus, this study uses both hypothesis testing and descriptive techniques to determine how pilots view their training for automation and the automation itself, as well as how and when they use it. In this way, general and specific threats to flight safety may be identified, and informed judgments regarding how to mitigate their effects can be developed.

## Background: The Systems View of Performance in Aviation

Despite the widespread use of automated aircraft, human error and accidents continue to occur, sometimes involving the most highly automated commercial aircraft. Investigations of both conventional and automated aircraft accidents in which crew error is implicated have shown that the majority of these errors involve interpersonal rather than technical deficiencies (Billings & Reynards, 1984; Cooper, White, & Lauber, 1979; Foushee & Manos, 1981; Gregorich, Helmreich & Wilhelm, 1990; Helmreich & Foushee, 1993; Murphy, 1980; Ruffell Smith, 1979; Wiener, 1993b).

In order to understand how these interpersonal deficiencies affect performance, it is helpful to conceptualize performance from a systems perspective. This perspective is predicated on a model that consists of inputs such as pilots' personalities, attitudes and skills, the physical and regulatory environment, and the organizational and national culture; group processes such as communications, decisions, and actions that play a part in task performance; and outcomes. Outcomes can be both proximal (i.e., mission and crew performance factors) and distal (i.e., individual and organizational factors [Helmreich & Foushee, 1993; Maurino, 1993]).

This model indicates that performance in aviation is team based – outcomes are dependent not only upon individual performance, but also upon how team members coordinate their activities and influence each other's process and outcome factors. This view of crew performance led to better understanding of a number of aviation accidents (e.g., National Transportation Safety Board [NTSB], 1979, 1982, 1994) which were caused by crew failures in interpersonal communications, task management, and group decision making (Helmreich & Foushee, 1993; Murphy, 1980). Cooper, White and Lauber (1979) categorized many of these failures as preoccupation with minor technical problems; inadequate leadership; failure to delegate tasks, assign responsibilities, and set priorities; inadequate monitoring and use of available information; and failure to communicate intent and plans.

However, good team performance is not seen as being perfectly free of errors. that experiments using pilots in high fidelity simulations show that pilots *typically* make an average of 3 to 10 errors per hour, but detect and recover from most of them (Amalberti, in press). Recently, Merritt and Helmreich (1996a; see also Merritt, 1996), noting that commission of error is unavoidable in any human endeavor, have sought to "normalize" (i.e., acknowledge and accept the inevitability of) error in this systems view, instead of punishing its commission. This paves the way for the design and implementation of safeguards and countermeasures against error. The goals of training and operational practices then become 1) reducing the likelihood of error, 2) 'trapping' errors before they have an operational effect, and 3) mitigating the consequences of error (Merritt & Helmreich, 1996a; Wiener, 1993a).

## Intervention Strategies for Managing Error in the System

Airlines and air crew training organizations have utilized this information described above to aid in the design of team-based training, commonly referred to as Cockpit or Crew Resource Management (CRM), also generically known as aviation human factors training. This type of training is defined by Lauber (1984) as "using all available resources – information, equipment, and people – to achieve safe and efficient flight operations" (p. 20). Edwards (1972) terms these resources, respectively, software (e.g., information, rules, laws, and operating procedures), hardware (e.g., equipment, instruments and materials), and 'liveware' (i.e., the people within the system). In this type of training, individuals are trained to enhance team performance by monitoring and verifying the performance of others comprising the team.

As a result of the continued application of sophisticated automation to flight deck tasks, many researchers and industry groups consider the FMS part of the air crew (ALPA, 1996; Billings, 1997; Woods, 1996). That is, because it performs strategic tasks (sometimes with a high degree of autonomy), it is seen as an "electronic crew member" that performs group process functions and interacts with live crew members (Helmreich, 1987, p. 70). While this view may risk anthropomorphizing flight deck automation, it is true that when used at high levels of automatic control, advanced flight deck automated systems can perform practically every function of human pilots (National Academy of Sciences Panel on Human Factors in Air Traffic Control [NAS], in press). Additionally, on highly automated flight decks many actions that are manually performed by pilots can be carried out solely through interface with the FMS. This understanding of automation strongly suggests that the manner in which crew members work with their "electronic peer" (Helmreich, Chidester, Foushee, Gregorich & Wilhelm, 1990, p. 13) has become similar to the manner in which crews work with each other. That is, for optimal performance, air crews must be trained to inquire into and verify the performance of automation. Furthermore, this becomes doubly important when the potential effects of automation use on both individual and crew performance are taken into account.

## Consequences and Potential Effects of Automation on the Operators

Automation of dynamic processes is associated with a host of other effects encountered by operators. These can generally be categorized as 1) complacency and overtrust; 2) excessive mistrust; 3) individual communication and workload modification; 4) degraded situation awareness; 5) loss of perceived authority and control; 6) loss of manual proficiency; and 7) team communication and coordination effects (Billings, 1997; Kantowitz & Campbell, 1996; Sherman, Helmreich, & Merritt, in press; Wickens, 1994).

**Complacency and overtrust.** Spurred by the occurrence of numerous transportation accidents attributed to operator failure in monitoring and cross-checking automated systems (NTSB, 1973; 1984; 1986; 1997), researchers have speculated that automation could lead to complacency and overtrust among operators. Until recently, empirical evidence for this was unavailable. Recently, Parasuraman and colleagues, in a series of laboratory studies using college students and professional pilots performing flight tasks in low-fidelity simulators, demonstrated that automation use can lead to complacency in monitoring and lessened awareness of automation failure (Parasuraman, Mouloua, Molloy, & Hilburn, 1993). This was especially so when it was very reliable (Parasuraman, Molloy, & Singh, 1993), and when it was used for extended periods (Hilburn, Molloy, Wong, & Parasuraman, 1993). Additionally, Riley (1996) determined, in a series of laboratory studies using college students and professional pilots, that low trust in automation delayed its use, but only in the early stages of experience with the system. Riley also demonstrated, in accord with Hilburn et al, that when automation was found to be highly reliable, subjects did not respond as quickly to initial indications of failure. This phenomenon was seen by the NTSB (1997) as causal in a recent maritime accident where the bridge crew did not notice the failure of a previously reliable automated navigation system, and ran aground 17 miles from their intended course.

**Excessive mistrust.** While a measure of mistrust of automation is probably associated with greater vigilance, excessive mistrust can lead to undesirable outcomes, such as underuse of appropriate automation and failure to realize its performance and economic benefits. Wickens (1994) identifies three possible sources for excessive mistrust: automation failures, perceived automation failures, and failure to understand certain aspects of automation function.

Trust in automated systems can be reduced when users encounter its failures (Muir, 1994). The second source of mistrust can result from erroneous starting input supplied by the operator, in which case the automation accurately performs functions using incorrect inputs. Automation can be perceived to have failed when it performs functions that the user does not expect to be performed. Wiener (1989), and Wiener, Chidester, Kanki, Palmer, Curry and Gregorich (1991), in an assessment of pilots' attitudes toward automated systems, found that commercial pilots sometimes expressed bewilderment and frustration with the behavior of certain computer-controlled aircraft systems, resulting in lowered trust in the systems.

The third source of mistrust described by Wickens stems from failure to understand the automation, which is seen as arising from an inaccurate mental model of the system. Sarter and Woods (1992, 1994) studied pilots' mental models of automated systems behavior on several automated aircraft, and found that many pilots held incorrect or incomplete notions concerning the behavior of various systems, especially in relation to operations involving automation-initiated changes to speed and navigational parameters.

In an elaboration of this work, Hansman and colleagues at NASA-Langley Research Center (Hansman, Midkiff, Vakil, Vaneck, Corwin, Irving, Irving, & Polson, 1995) recently showed that whereas most flight deck automation is dependent upon multiple inputs and can simultaneously effect changes to parameters in multiple subsystems, pilots' conceptions of systems behavior usually represent the systems as functioning in a single-input-single-output mode. In other words, while an automated system execute many actions at once, pilots tend to view the system as executing actions serially. Perhaps as a result of these biases, it has been shown that pilots typically tend to use only a subset of a given system's available modes (Degani, Shafto, & Kirlik, in press).

Incorrect or incomplete conceptions of automation are thought to have been responsible for at least five aircraft incidents and accidents over the past seven years (Dornheim, 1995). Two strikingly similar incidents in which pilots became confused regarding system modes have occurred (both involving the Airbus A320, one of the most highly automated commercial aircraft), one over San Diego in 1990, and another on approach to Strasbourg, France in 1992. In both situations, evidence suggests that the pilots were operating under the assumption that the aircraft was in one type of approach mode ("angle-of-descent" mode), when in fact the aircraft was in an approach mode with different input requirements ("vertical speed" mode). In both cases, pilots inadvertently entered vertical speed instead of angle-of-descent data, with the result that the aircraft began descending at a rate of several thousand feet per minute. One aircraft was recovered by the pilots after a brief struggle. The other aircraft crashed short of the airfield, resulting in many fatalities (Dornheim, 1995).

**Individual workload modification.** With the introduction of automation, workers who formerly monitored their own and others' performance are charged with the task of monitoring both their own and others' performance *and* the automation, and adequately communicating the state of the operation to one another (Wiener & Curry, 1980). This requires greater expenditure of attentional resources (Pew, 1969; Sheridan & Johannsen, 1976). Kantowitz and Casper (1988) describe how automation use can affect workload:

> Tasks once performed by the crew can now be allocated to automatic devices; this is intended by designers to reduce crew workload. However, the crew is still responsible for the operation of the automatic device, and this introduces monitoring requirements that increase workload (p. 161).

Evidence for this is provided by Sarter and Woods (1992), who, in a survey of 135 automated aircraft pilots, found that monitoring and tracking systems performance required more cognitive workload in highly automated aircraft than in less automated aircraft.

**Degraded situation awareness.** While the integration of critical information afforded by automated systems (e.g., color radar showing the location and intensity of storms, superimposed on a computer-generated map display) can bolster crews' situation awareness, automation use can also degrade it. Situation awareness is generally defined as an individual's overall sense of an operation's past and probable future condition (Wickens, 1994). Pilots of highly automated aircraft often need to direct their attention inside the cockpit in order to monitor the systems, causing a diversion of their attention away from the environment in which they are maneuvering.

In general, situation awareness can be lowered when using automation because the operator, instead of performing control actions and receiving on-line feedback regarding operational conditions, must monitor the automation's performance of control actions and extract information regarding conditions without direct feedback (such as that gathered from hands-on command of aircraft controls). It has been demonstrated that when operators are passive monitors of other's input, as opposed to active suppliers of input, it becomes more difficult to understand, learn, and remember consequences of the inputs (Endsley & Kiris, 1994).

**Loss of perceived authority and control.** According to Wickens (1994), the efficiency of automation and the incorporation of performance limitations that prevent operators from exceeding the structural limits of the hardware can cause an operator to perceive a loss of control and authority. Automation can also erode an operator's perceived authority if the systems begin performing actions that *appear* to be outside the operators' ability to control.

The cockpit voice recorder recovered from the April 1994 crash of a China Air A300-600, in which the pilots struggled with the autopilot for control of the aircraft, indicates that loss of perceived and actual authority may have played a role in this accident. In the final seconds of the flight, the captain attempted to disconnect the autopilot using manual force on the control stick. However, this only partially disconnected the autopilot, and the crew was unable to achieve their objective, which was to continue the approach to the runway. The flight management system then initiated control actions to countermand the crew's control actions. As the aircraft continued to pitch up, the captain exclaimed "Goddamn it, why it comes in this way?", referring to the automated system actions. Soon after, the airplane stalled in a 50-degree nose-up position and crashed. Throughout this episode, the crew had the ability to disengage the automation by pushing the autopilot disconnect button; however, the crew continued to struggle against the autopilot instead of disengaging it (Dornheim, 1995).

**Loss of manual proficiency.** Automation, by definition, removes the necessity for direct human control of certain processes. Because the human operators are no longer performing control functions on a regular basis, their skill at performing these actions can atrophy (Bainbridge, 1983; Wiener, 1988). Wiener (1989), in a survey of approximately 200 pilots flying the highly automated Boeing 757, found that almost 90% reported manually flying for a portion of the flight, in order to preserve basic flying proficiency. Almost half of the pilots expressed concern about loss of manual flying skills.

Wiener and colleagues (Wiener, Chidester, Kanki, Palmer, Curry, & Gregorich, 1991) also assessed attitudes toward automation on the flight deck among a sample of Douglas DC-9 and McDonnell-Douglas MD-88 (an automated version of the DC-9) crews. Over 75% of respondents agreed with the statements "I take active measures to prevent a loss of my flying skills due to too much automation," and "In a highly automated plane, you run the risk of loss of basic flying skills."

McClumpha, James, Green and Belyavin (1991) conducted a study of attitudes toward flight deck automation among pilots in the U.K., using an attitude survey which assessed attitudes in several content areas, including workload, crew interaction, feedback, training, specific aspects of flying skills, and general evaluations of automation. Of the 572 pilots responding to the survey, most reported that their overall level of airmanship had been enhanced by automation, but that their basic manual skills had been eroded. Complementing these findings, Sherman, Helmreich, and Merritt (in press) found that up to 85% of pilots reported disengaging automation during some portion of flight in order to preserve manual flying skills, although this proportion differed considerably among pilots from different nations.

**Team communication and coordination effects.** The phenomena described above are mostly individual effects, although they can influence team performance. Automation can also affect group processes directly, by affecting the quality and quantity of human-to-human interaction in team endeavors. As mentioned earlier, if an FMS is provided with inaccurate yet plausible raw information (e.g., a transposed digit in a heading or altitude input), incorrect information can be propagated throughout the system.

Danaher (1980) points out that the intangible benefits stemming from the interaction of humans in a complex system (i.e., the opportunity for richer, more flexible information transfer) are reduced when the interaction is limited to the human-computer interface. This is implicitly recognized by at least some FAA inspectors, who recently reported that pilots at one major U.S. airline do not practice adequate verification and cross-monitoring of one another's automated system inputs (Dornheim, 1996a). Generally speaking, if crews do not verbalize and confirm their inputs to an automated system, then 'set-up' errors can turn into performance errors.

Hulin (1989) observes that when automated systems fail, the operators frequently need to increase their communication and workload:

> ...failures of automated sub-systems in group-operated technical systems create intensely information-dependent situations...the information required to diagnose and correct failures may be dispersed throughout the crew, and the necessity to use information available from other crew members means that intra-crew communication becomes crucial following failures of certain components of the overall systems (p. 775).

These observations suggest that operators will frequently need to increase levels of communication and coordination when utilizing automation in order to ensure safe accomplishment of tasks. This has been found in both simulator-based and laboratory studies. Using high-fidelity simulators and scripted flight scenarios, Veinott and Irwin (1993) found in a comparison of crews flying DC9 and MD88 aircraft that crews in the latter airplane communicated more, showing higher frequencies of commands, observations, questions, replies, and acknowledgments. Bowers, Deaton, Oser, Prince and Kolb (1993) reported an increase in overall communications for subjects operating simulated automated aircraft.

## Training for Automation Use in Aviation

Initially, as airlines took delivery of highly automated aircraft, aircrews transitioning to these aircraft were taught to use automation as much as possible, presumably because this would prevent error and increase operating efficiency (Hopkins, 1993; Laming, 1993). Soon, however, it became apparent to pilots and training departments that some of the above-mentioned effects (i.e., complacency, communication and coordination effects, degraded situation awareness, loss of proficiency, etc.) could be observed in actual line performance (Helmreich, Hines, & Wilhelm, 1996).

Attempting to counteract these deleterious effects, training departments at a few airlines formulated and disseminated guidelines (or ‘philosophies’) for automation management (Degani & Wiener, 1994). Generally, these automation use philosophies stressed the air crew’s responsibility remain proficient and comfortable with all levels of automation, the presumption being that remaining proficient in all aspects of automation use would prevent some of these negative consequences. Some training departments also began to integrate these philosophies of automation management into existing human factors (i.e., CRM/HFS) training, reasoning that automation management was more effectively presented in conjunction with training for management of other resources. Several years ago Delta Airlines implemented a training course called “Introduction to Aircraft Automation” ($IA^2$); this course encourages aircrews to remain proficient in the use of all levels of automation, and to exercise judgment in utilizing the degree of automation most appropriate for a given situation (Byrnes & Black, 1993). This approach is supported by findings from Riley (1996), who claims that encouraging rational use strategies through training is an effective means of changing aircrews’ behavior.

However, this type of training may not be as effective in ameliorating certain important operator-automation problems, such as inaccurate mental models of automation functions. As Sarter and Woods (1994), Billings (1997), and others point out, pilots in lab experiments and in line operations frequently do not understand how certain automated systems operate and/or why they are designed to operate the way they do. Their solution is to call for more training to explain “how the system operates and why, rather than simply how to operate the system” (Billings, 1997; p. 192). Some disagree with this position, citing the economic infeasibility of increasing the length and breadth of training (the ‘training footprint’), and observing that a pilot should not and does not need to know how an aircraft is built in order to safely fly it (Bent, 1996). Bent also points out that attempting to apply complex technical systems knowledge in emergencies could very well lead to errors that could worsen the situation (this concept is illustrated by the pilot-in-training’s lament “how am I supposed to remember this sh*t when I’m scared?!?”).

Recently, an approach that seeks to unite amelioration of deleterious operator-automation effects with an economical approach to training for use of automated systems has been implemented by American Airlines (AAL). At the end of 1996, American required that all pilots attend the Advanced Aircraft Maneuvering Program (AAMP). This single-day training session explicitly teaches that 1) automation should not be trusted implicitly; 2) if the crew does not understand what the automation is doing and/or why it is doing it, the automation must be disengaged and the aircraft operated manually; 3) if situation awareness and flight path management are compromised by the need to attend excessively to the automation, the automation must be disengaged and the aircraft operated manually; and 4) if crews feel that their manual flight skills need maintenance, they should revert to basic, non-automated flying (AAL, 1996). While these efforts to prevent negative consequences of automaton use are admirable, it has been observed that without pervasive organization-wide support, initiatives like these often meet with only limited success (ATA, in press; Degani & Wiener, 1994). Generally, these observations are based in the systems view described earlier, as they take the position that multiple factors affect flight crew performance.

It is important to note that these automation-specific initiatives have been accompanied by an FAA-led transition of fleet training programs from traditional, time-based training requirements to more individualized, proficiency-based training requirements. The new requirements are referred to as the Advanced Qualification Program ([AQP]; FAA, 1991). Qualifying a training program under this set of regulations requires comprehensive reporting and tracking of training quality; while these tasks take up considerable resources, many AQP proponents claim that training organizations can potentially realize substantial economic and safety benefits by moving to AQP-based training standards.

In summary, it is fair to say that the philosophies and practices of training for automation management are in a state of transformation. The FAA Human Factors Team (1996), in its report "The Interfaces Between Flightcrews and Modern Flight Deck Systems", recommends that all automated aircraft flight training programs focus on imparting the types of lessons taught in $IA^2$ and AAMP, using a variety of training methods. In particular, they recommend using as teaching tools incidents and occurrences involving automation that have been experienced by pilots, as well as descriptions of pilots' "mental models" (p. 94) of automation. They also recommend that airlines support these initiatives throughout the organizational structure, not just in the training department.

In order to ensure training efficacy, both human factors and automation management training require rigorous design and evaluation. Survey research methods have proven particularly useful in identifying potential threats to safety that are addressable in training, and in determining whether training has resulted in desired attitude (and by implication, behavioral) change. Although some researchers call into question the utility of survey research, when performed correctly, this method can provide clear indications of probable behavior, especially when both the attitude and the behavior are specific and closely related (Ajzen & Fishbein, 1977; Fazio, 1986; Fazio & Zanna, 1981).

To design and implement more effective, empirically-informed training philosophies and practices (i.e., in order to better address potential threats to safety with training), it was recognized that more information was needed regarding aircrews' evaluations of their own and the automation's performance, their evaluation of the training they received, and their attitudes toward automation use in more specific situations. To accomplish this, a detailed survey instrument was created that would capture this type of information, as well as information regarding air crews' evaluations of the automation on their aircraft.

## THE UNIVERSITY OF TEXAS AVIATION AUTOMATION SURVEY

The survey tool, entitled the University of Texas Aviation Automation Survey (AAS), was created to allow further measurement of pilots' responses to FMAQ automation items, and also allow investigation of pilots' views of the usability of aircraft-specific systems such as the FMS, CDU, electronic flight instrument systems (EFIS), and autothrottles, as well as the relationship of pilots' training background to their evaluation of these systems. It consists of five sections. The instrument appears in Appendix A.

Section I contains demographic items that record respondents' nationality, nationality at birth (if different than present nationality), flying background (military or civilian), total years piloting, total flight hours, current aircraft, flight hours in current aircraft, experience and flight hours in previously flown aircraft, crew position (captain/pilot in command, first officer, or relief first officer), and whether the respondent has other job responsibilities (e.g., instructor pilot, management pilot, etc).

Sections II and III requests information concerning respondents' training experience. It probes whether pilots were allowed 'free-play' (i.e., unsupervised, discretionary practice time) on a part-task simulator or CDU during training. The training evaluation items also prompt for respondents' evaluations of the quality and comprehensiveness of initial and recurrent training, the usefulness of abnormal procedures, and computer-based training, and overall assessment of the applicability of basic aviation skills to the tasks required in their present aircraft.

The remainder of Section III contains a pool of items probing pilots' evaluations of the various automated equipment on the aircraft they currently fly, as well as items probing their evaluations of the training they received during transition to this aircraft. The equipment items inquire about the flight management systems' ease of use, reliability, input efficacy, outputs and feedback efficacy, and the usefulness of FMS-issued error or diagnostic messages on their current aircraft. Items in several of these domains are conceptually based on items from McClumpha, James, Green and Belyavin's (1991) survey of United Kingdom pilots flying automated aircraft. The training items inquire whether respondents feel that transition training was adequate preparation for flying their current aircraft in both normal and emergency situations, and whether basic flying skills are useful in their current aircraft.

Section IV contains the pool of twenty automation items from the Flight Management Attitudes Questionnaire (FMAQ; Helmreich, Merritt, Sherman, Gregorich, & Wiener, 1993). The FMAQ measures attitudes toward specific domains of interpersonal performance on the flight deck, and use of automation in a team environment. Section V contains items from FMAQ scales derived by Merritt (1996), including the Uncertainty Avoidance Values scale. Uncertainty Avoidance is seen as a relatively enduring personal characteristic that reflects a need for reduced ambiguity in many aspects of life, including work. High need to avoid uncertainty can manifest itself in the form of increased preference for proceduralization and of work tasks and formalization of rules and codes for behavior.

Using the AAS, the following theoretical expectations regarding perceptions of training, evaluations of equipment, and attitudes toward automation management were explored among a sample of U.S. pilots presently flying automated aircraft.

## THEORETICAL EXPECTATIONS FOR EVALUATIONS OF TRAINING, EQUIPMENT AND AUTOMATION MANAGEMENT

### Pilots' evaluations of training

**Relationships with organizational variables and aircraft type.** The FAA states that one potential threat to safety is the reduced ability of air crews to understand and respond appropriately to "non-normal circumstances" or "novel malfunctions" when flying highly automated aircraft (1996; p. 81). They also observe that, generally, air carriers differentially prepare crews to cope with "probable but unusual situations" in training, and recommend greater focus on this in flight training. Since training programs at different airlines are typically a mixture of industry-wide techniques, manufacturer recommendations, and in-house (i.e., idiosyncratic) procedures and traditions, it is expected that pilots' evaluations of training will differ across organizations. Similarly, as aircraft fleets have different management teams and training requirements, it is also expected that evaluations of training will differ across fleets, both across and within organizations.

It should be noted that directional inferences for these hypothesized effects are beyond the scope of the present study. This is so for two reasons. One, it could be considered disingenuous to speculate on the meaning of cross-fleet and cross-organization differences for ratings of training efficacy, without accounting for contextual information regarding the training programs being rated. Two, putting any observed differences into sufficient context would risk identifying the airlines from which the pilots' responses were drawn.

**Relationships with pilot experience.** An important question that *can* be addressed in the present study is whether and how characteristics of individual pilots are associated with reports of training efficacy. Not only can training programs differ, the lessons learned by trainees can differ, even when trainees undergo the same training program. For example, pilots with previous experience in automated aircraft may already possess some of the requisite skills and knowledge to enable them to anticipate and deal with probable but unusual situations when using automation. Pilots without such experience may lack the skills and knowledge to cope with these types of situations.

The FAA concludes that pilots with no previous automation experience are most vulnerable to inadequate or inaccurate understanding of automation, and have explicitly recommended that “training should be adapted to the background of the pilot (e.g., glass[3] vs. non glass experience)” (1996; p. 110). Following the FAA’s line of reasoning, it could be predicted that pilots with no previous experience flying automated aircraft will feel less prepared for line flying after transition training, report developing a lesser understanding of the FMS during training, and feel that they learned more about the aircraft in actual line flying than in training, relative to pilots with previous automated aircraft experience.

Alternatively, it is possible that mere exposure to the tasks of flying (either automated or non-automated aircraft) would prepare pilots for operating an automated aircraft. In this view, pilots’ reported understanding of the FMS and preparation for line flying would be positively related to their overall experience level (as measured by total flight hours).

**Relationships with the availability of ‘free-play’.** During training, pilots invariably follow a fairly rigid curriculum. However, most training departments recognize that pilots who are transitioning to an automated aircraft are simultaneously highly motivated to succeed and sometimes overwhelmed by the complexity of the automated equipment. Therefore, it is common practice to offer pilots the opportunity for non-jeopardy, individual, discretionary practice on various training devices, including part-task devices designed to develop FMS skills and knowledge. This component of training is widely known as ‘free-play’. A recent study indicates that about 80% of pilots avail themselves of free-play opportunities during training (Sherman & Helmreich, in press). Furthermore, a large percentage of pilots find this training quite useful. Assuming that no negative learning occurs as a result of effective free-play opportunities, it is reasonable to posit that pilots who availed themselves of free-play and found it useful will feel more prepared for line flying after transition training, report developing a greater understanding of the FMS during training, relative to pilots who didn’t find free-play useful or who didn’t avail themselves of free-play.

**Pilots’ evaluations of automated equipment**

The FAA (1996) also recommends further study of how the automation on different aircraft is perceived by pilots, as a means of further understanding crew-automation interaction. This section is devoted to theoretical expectations concerning pilots’ evaluations of automated equipment.

**Relationships with aircraft type.** As discussed earlier, other investigations have shown that pilots’ evaluations of automation seem to differ substantially across aircraft types (Rudisill, 1994; Sarter, 1996; Sarter & Woods, 1995). Taking an overview of the polemic surrounding flight deck automation, two opposing perspectives concerning the increasing automation of flight tasks can be distinguished. One outlook holds that early (circa 1980’s) implementations of flight management systems were balky and user-hostile (i.e., ‘clumsy automation’ [Wiener & Curry, 1980]); although with increasing experience and technological innovation, airframe manufacturers are building better, more user-friendly flight management systems (e.g., Bent, 1996). This outlook is termed the ‘utopian view’, as it implies that automation implementations are generally moving toward a brighter, better future. Proponents of this view, therefore, would hold that pilots flying newer-generation automated aircraft will report that their aircraft’s FMS is easier to use and understand, and the aircraft promotes instead of degrades pilots’ awareness of both systems states and the overall operation.

---

[3] “Glass” and “non-glass” are colloquialisms for automated and standard aircraft, respectively. The term “glass” derives from the fact that most (but not all) FMS-equipped aircraft use large CRTs instead of small mechanical gauges for the primary flight displays. Thus, the instrument panels on these aircraft consist of large expanses of glass screens instead of isolated round-dial instruments.

A contrasting outlook holds that since the early 1980's, aircraft manufacturers have added features and functionality to the point that flight management systems have become increasingly (and unnecessarily) complex, opaque, and difficult to use (c.f. ALPA, 1996, FAA, 1996). This outlook is termed the 'dystopian view'; it implies that as technology marches on, flight management systems will become increasingly more complicated and difficult to use. Proponents of this view would assert that pilots flying newer-generation automated aircraft will report that their aircraft's FMS is more difficult to use and understand, and the aircraft degrades instead of promotes awareness of both systems states and the overall operation.

An alternative to these two views is that because the flight deck configurations of the various automated aircraft are so different, generation-related generalizations are not justifiable. If this view were tenable, then there would be no discernable generation pattern to equipment ratings across aircraft types; instead, each aircraft would have it's own unique configuration of ratings.

**Relationships with training and previous experience.** Previous research (e.g., Amalberti, in press; Sherman & Helmreich, in press) indicates that a pilot's initial experience with an automated aircraft after training can be accompanied by difficulties in understanding the aircraft's FMS. Given that training can differ considerably across airlines and even fleets, it is likely that different training programs will have varying levels of success in fostering FMS understanding. It is possible that these differential levels of training efficacy could be associated with evaluations of automation. A reasonable supposition, then, would be that evaluations of their present aircraft's FMS are positively related to pilots' perceptions of training efficacy.

Amalberti also states that a sufficient level of comfort, understanding and proficiency with the FMS can elude pilots even after 1,000 flight hours (i.e., more than a year) in their current aircraft. This is probably due to the fact that, as stated by David Woods, effective use of automation often requires the development of "new knowledge and skills" (FAA, 1996; p. 86) on the part of users. Even among pilots who are transitioning from one automated aircraft to another, or who have prior experience with automated aircraft, the first year on a new aircraft can be characterized by much trial-and-error learning.

These observations suggest that pilots' self-reported proficiency and comfort with their aircraft's automation may be related to their previous aircraft experience and/or experience in their current aircraft. As described above, it may be reasonable to expect that pilots with no previous flight deck automation experience are more vulnerable to inadequate or inaccurate understanding of automation than are pilots with previous experience. Thus, it would follow that pilots with previous experience in automated aircraft will report that their current aircraft's FMS is easy to use and understand, relative to pilots with no previous experience.

Unlike the training item section, which prompts for evaluations that are rooted in the psychological past (e.g., "I gained an adequate understanding of the FMS during transition training"), the equipment items prompt for evaluations that are rooted in the psychological present ("The FMS on this aircraft is easy to use"). It is therefore reasonable to expect that pilots' reported comfort and proficiency might positively covary with experience in their current aircraft (as measured by flight hours in their current aircraft). Following Amalberti's and Wood's reasoning, it can be hypothesized that as pilots gain more experience in their current aircraft, they will report that their aircraft's FMS is easier to use and understand.

There is another possible demographic source of variation in automation evaluations. As described in the previous section on training, mere exposure to the tasks of flying (either automated or non-automated aircraft) could theoretically be associated with greater understanding of the FMS on automated aircraft. Thus, pilots' positive evaluations of automation and their total years of flying experience would be positively related. Alternately, as Wiener (1988) suggests, younger, less experienced pilots (who on average probably have had more exposure to computers than their older counterparts) may be more comfortable with the logic and behavior of computerized systems in general. This comfort may extend to computerized flight guidance and control. If this were so, then it would be reasonable to expect that evaluations of automation and pilots' total years of flying experience would be inversely related.

At this point, it should be pointed out that the items in the equipment section of the AAS prompt for very specific, bounded evaluations of the various aspects of aircrafts' flight management systems. Therefore, in analyses across aircraft types, pilots' ratings for different aircraft will be presented; i.e., aircraft type will *not* be de-identified.

**Pilots' automation management attitudes**

This section describes some theoretical expectations for pilots' self-reported management of automation and management of flight deck tasks when using automation (for expediency, these concepts are referred to as "automation management attitudes"). In order to develop this line of reasoning it is useful to once again return to the FAA's 1996 report. In this report the FAA observed that many carriers lack an automation operating 'philosophy' (i.e., an overarching set of guidelines for the effective, safe use of automation) that takes into account the potentially deleterious effects of automation use. As mentioned earlier, those carriers that have promulgated an automation use philosophy generally have not successfully communicated it to pilots (ATA, in press). The FAA (1996) states that the articulation and dissemination of an empirically-informed philosophy of automation use is a particularly important means of reducing threats to safety, and explicitly recommends that carriers ensure training covers appropriate "management and use of automation." (p. 109).

It is highly likely that factors other than training can affect pilots' self-reported automation management. It has been demonstrated that national variability is associated with differences in automation management attitudes, both within and outside aviation (Forslin, Soderlund, & Zackrisson, 1979; Sherman, Helmreich, & Merritt, in press). Sherman, Helmreich, and Merritt also demonstrated very small cross-organizational differences within each of several nations included in the study. It is possible (but not likely) that automation management attitude differences will be found among the organizations represented in the present study. A finding of differences in pilots' automation management attitudes across organizations would be intrinsically interesting; however, the mere existence of variations in responses alone does not aid in the amelioration of safety threats[4]. There are, however other potential sources of variation (both individual- and organization-based) that may prove useful in the ameliorating threats to safety.

**Relationships with pilot experience.** In previous sections, the position is taken that pilots with no previous automation experience who are transitioning to an automated aircraft might be most vulnerable to inadequate or inaccurate understanding of automation. This is because, theoretically, these pilots have no relevant experience base from which to assimilate their new knowledge and skills. Following this line of reasoning, it is feasible that pilots with no previous automation experience would also possess lesser awareness of the potentially deleterious effects of automation use (such as degradation of manual skills, degraded vigilance, increased communication and coordination needs, etc).

Pilots' recognition of these effects may increase with experience in automated aircraft. Therefore, it can be predicted that pilots with previous experience in automated aircraft will show greater recognition of the potentially deleterious effects of automation use. If this were so, it would imply that training organizations should be particularly careful to ensure their newest trainees leave the schoolhouse with an appreciation of automation's limitations and consequences. Alternatively, it is also possible that pilots with previous experience using (typically very reliable) flight deck automation may discount the risks and limitations of automation. Therefore, it could be hypothesized that pilots with no previous experience in automated aircraft will show greater recognition of its potentially deleterious effects. If this were so, it would imply that training departments need to ensure that pilots with previous experience in automated aircraft do not become complacent in their management of the automation.

**Relationships with measures of organizational policy**. As described in the introduction, aviation can be considered from a systems standpoint. In this view, organizational philosophies, policies, and procedures are all seen as determinants of pilots' attitudes, values, and actual practices. As mentioned earlier, most carriers do not explicitly promulgate a comprehensive philosophy and policy regarding automation use, and those that do are only partially successful at disseminating the information. However, it would be a mistake to assume that pilots within an airline

[4] However, the measurement and description of response variation *within* a particular airline may be helpful to an individual airline.

with no explicit automation guidelines do not *perceive* (through the transmission of organizational norms) some type of implicit organizational inclination regarding automation. Therefore, it could be expected that pilots who perceived that their company favored maximal automation use would endorse less discretionary use of automation.

**Relationships with measures of uncertainty avoidance.** Those with a high level of uncertainty avoidance tend toward a preference for greater formalization of rules and codes for behavior, and increased proceduralization of work activities (Hofstede, 1980), as this aids in reducing ambiguity and increasing prediction and control of future events. Merritt sees this construct as a relatively enduring personal characteristic (1996). Use of automation to control flight and navigation tasks can reduce *perceived* (but not necessarily *actual*) ambiguity in flight control and flight management. Because uncertain, ambiguous situations are seen as an unpleasant, motivating state for individuals with high uncertainty avoidance, pilots with this characteristic should tend toward greater reliance on automation as a means of reducing uncertainty in the performance of job tasks. They should also prefer automation more than do pilots with low uncertainty avoidance needs.

This section has outlined many expectations for relationships between demographic and systemic variables, and measures of pilots' training evaluations, equipment evaluations, and attitudes toward automation management. In later sections, the relationships explicated here will be explored in analyses. Before analyses are presented, information regarding the acquisition and composition of the sample is provided.

## SAMPLING AND SAMPLE DEMOGRAPHICS

### Sampling strategy

The AAS was administered during the winter of 1996-1997. A total of 1,718 surveys were collected. Three groups sponsored the survey administration. The Air Line Pilots Association (ALPA), the largest U.S. commercial pilot union, administered the survey to a group of their pilots who fly automated aircraft; the flight operations departments of two major airlines also sponsored survey administration among their pilots.

Through an agreement with ALPA, the AAS was administered in February, 1997 to 4,500 pilots, all of whom presently fly automated aircraft. ALPA survey administration took the form of a simple random sample. As a result, the final tally by airline and aircraft type is representative of the airlines' pilot populations and fleet sizes. That is, relatively more surveys were returned by pilots from larger airlines and relatively fewer by pilots from smaller airlines, and relatively more surveys were returned from the automated larger fleets within each airline (for those airlines represented in the sample by more than one automated fleet). ALPA pilots returned a total of 1,410 surveys. This corresponds to a response rate of 31%, which Kalton (1983) defines as a slightly low but still acceptable response rate for mailed surveys.

A total of 308 surveys were collected directly from two airlines during November 1996 and January 1997, as part of unrelated projects measuring the performance of flight crews during line operations. At one airline, pilots were handed the survey by a flight deck observer at the conclusion of the observed flight (or flights, if the crew was observed for more than one flight segment); approximately 450 surveys were administered in this fashion. The second airline included the survey in pilots' monthly bid packets for one automated fleet. This administration surveyed the entire population of pilots in this fleet at this airline (N ~ 400). The airline that distributed the survey by hand contributed 159 surveys and the airline that mailed the survey contributed 149 surveys, representing response rates of 35% and 37%, respectively[5].

### Respondent demographics

The respondents represented ten U.S. airlines, although only seven of the ten were represented by sufficient numbers of respondents to warrant inclusion in cross-airline comparisons (the three excluded airlines contributed only five respondents total; these respondents are included in analyses where airline is not utilized as a factor). Airline 1 contributed 159 respondents; Airline 2, 96; Airline 3, 159; Airline 4, 409, Airline 5, 149; Airline 6, 644;

---

[5] These two airlines received a proprietary report detailing their pilots' responses to the survey.

and Airline 7, 95 (Airlines 1 and 5 are the carriers where the surveys were administered by the flight operations departments)[6]. In keeping with the University of Texas Aerospace Crew Research Project's data collection agreement with the participating airlines, the respondents' airlines have been de-identified. To preserve the anonymity of the airlines and provide an adequate description of the sample, only the *number* of automated fleets representing each airline in the sample, up to a limit of three, are described. Airlines represented in the sample by three or more automated fleets will only be identified in the sample as represented by three or more fleets; i.e., the exact number of fleets will not be given when it exceeds three. It should be noted that the fleet number for each airline may or may not correspond to the total number of automated fleets that are actually *operated* by each airline; this provides an added level of anonymity. Responses from three or more automated aircraft fleets were received from Airlines 1, 4, 6; responses from two automated fleets were received from both Airline 3 and Airline 7; Airlines 2 and 5 are each represented by responses from only one automated fleet.

For selected analyses it is possible to compare responses across two widely operated aircraft types, the B737-300 and B757/767. Table 1 contains the total number of respondents from each of the seven airlines, as well as the number of fleets representing each airline, and the number of responses (if any) for each airline from the B737-300 and the B757/767 fleets.

[6] There were *two* airlines with precisely 159 respondents.

Table 1. Total Number of Respondents from Seven Airlines, Number of Fleets Representing Each Airline, and the Number of Responses (if any) for Each Airline From the B737-300 and the B757/767.

| | Airline | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Subject *n* | 159 | 96 | 159 | 409 | 149 | 644 | 95 |
| Fleet *n* | $\geq$3 | 1 | 2 | $\geq$3 | 1 | $\geq$3 | 2 |
| B737-300 | --- | 95 | --- | --- | --- | 176 | 49 |
| B757/767 | 62 | --- | 95 | 89 | 149 | 95 | 46 |

*Note.* Airlines without entries for B737-300 and B757/767 rows are not represented in the sample by these aircraft.

Nine different automated aircraft types were represented by the respondents. Only seven different types of aircraft were represented by sufficient numbers of respondents to include in cross-aircraft type analyses. The A320 was flown by 438 respondents; the B737-300 and later series, 325; the B747-400, 145; the B757/767, 537; the B777[7], 105; the MD11, 60; and the MD80 and later series, 66. Seven of these aircraft types were represented by sufficient numbers of respondents to warrant inclusion in cross-aircraft type (i.e., fleet) analyses.

The majority of the respondents (1,437) were regular line pilots; the rest were check pilots (166), instructor pilots (64), and management pilots (11). Forty respondents did not indicate their status. A little over half (828) of respondents who indicated their initial flight training background were initially trained in the military; 713 respondents indicated their initial training was as a civilian, and 177 did not indicate their flight training background. A total of 1,413 reported being able to practice on a part-task FMS training device while training for their current aircraft; 247 did not practice on this type of device (31 respondents did not answer this question).

Of the 1,718 respondents, 1,668 are male and 37 are female. Thirteen respondents did not report their gender. A total of 1,701 reported their nationality as American; the rest were Canadian (3), Austrian (1), British (1), Chinese (1), French (1), German (1), Italian (1), and Norwegian (1). Six respondents did not report their nationality. It has been shown that nationality is significantly associated with pilots' attitudes and evaluations of automation in aviation (Sherman, Helmreich, & Merritt, in press), therefore the responses from non-U.S. and unidentified nationals were not included in analyses.

The respondents reported a wide range of experience. Total years in aviation for these respondents ranged from 0.3 to 50 years[8]. The median reported years in aviation was 22. Total flight hours ranged from 1,100 to 36,000 (median: 14,000). Respondents' total years with their organization ranged from 0.1 to 40 (median: 15). Respondents' flight hours in their current aircraft ranged from zero to 22,180 (median: 1,500). As expected, junior officers (i.e.,

[7] It is common knowledge that the B777 is operated by only one U.S.-based carrier. Although identifying this fleet by name provides a de facto identification of this airline, the potential for information to be gained by fleet-level comparison of the state-of-the-art B777 with other automated aircraft was worth the price of identifying this airline. To avoid providing a full identification of this airline's data, airline codes (i.e., Airline 1, 2, 3, etc) will not be linked to the B777 fleet in any tables or analyses.

[8] As this is an abnormally high number, this respondent's supporting information was inspected. This respondent had amassed nearly 30,000 flight hours (commercial pilots usually average about 750-800 hours per year). Most likely, this individual served as instructor pilot after reaching age 60, or this datapoint was misrecorded.

first officers) had less experience on all these demographic indicators. Experience levels differed at each airline; median total flight hours across airlines ranged from 10,000 (Airline 1) to 20,000 (Airline 3)[9].

There was a preponderance of senior officers in the sample; 1,193 (69%) reported that they were captains, and 518 (31%) reported that they were first officers. Seven respondents did not report their position. The overrepresentation of captains in the sample initially suggested a response bias; however, the ratio of captains to first officers from the airline-administered groups (Airlines 1 and 5) was almost exactly 1:1. This suggests that the uneven ratio was yet another artifact of the ALPA sampling strategy, and not a systematic respondent bias.

To summarize the characteristics of the sample: 1,718 surveys were collected. After eliminating non-U.S. and unidentified nationals, as well as those who did not fly automated aircraft or did not indicate what type of aircraft they flew, a total of 1,691 were deemed appropriate for use in subsequent analyses.

## METHOD AND RESULTS FOR EVALUATIONS OF TRAINING

This section describes the analytic methods used to investigate the training item included in the survey. It also presents the results of the analyses. Additionally, tertiary analyses concerning the effect of AQP curriculum application are provided at the end of this section.

### Analytic method

**Independent variables and covariates**. Earlier it was posited that multiple factors may influence pilots' evaluations of their transition training – these factors are the airline for which they work (i.e., organizational culture), their fleet (i.e., aircraft type), their previous experience (if any) with automated aircraft, and evaluations of the usefulness of discretionary part-task training. These variables are utilized as independent variables in the multivariate analysis of covariance (MANCOVA) and analyses of covariance (ANCOVAs) described below[10].

Also, it was suggested that pilots' overall experience could be related to their evaluations of training. Because experience as measured by the survey is a continuous, not a discrete, variable, it is utilized as a covariate, not as levels of a factor in an analysis of variance. However, it is recognized that the presence of a significant relationship between experience and training evaluations would have meaningful implications in its own right; therefore, experience effects will be analyzed and discussed independently.

**Dependent variables**. Of the items concerned with training, six have an explicit evaluative component. The six items are "My transition training adequately prepared me for flying this aircraft on the line", "Recurrent training for my aircraft covers important abnormal procedures that are not encountered very often", "I gained an adequate understanding of the FMS during transition training", "I learned most of what I need to know about this aircraft when I started flying it on the line", "My company's computer-based training for this aircraft is effective", and "If [you were able to practice on a part-task FMS device], how useful did you find this type of training?". Additionally, one item (item 32, "During transition training, I was taught that the decision to use or not use automation on this aircraft is up to me") reflects whether pilots felt that discretionary (as opposed to maximal) automation use was stressed during transition training.

Three of these items ("My transition training adequately prepared me for flying this aircraft on the line", "I gained an adequate understanding of the FMS during transition training", "I learned most of what I need to know about this aircraft when I started flying it on the line") form a conceptual grouping – they elicit pilots' evaluations of

---

[9] Those airlines with relatively lower overall median flight hours (Airlines 1 and 4) had recently begun an intensive hiring push. Airlines with high overall median flight hours have not hired many pilots recently.

[10] In addition to the statistical assumptions inherent to ANOVAs and MANOVAs, analysis of covariance requires that the assumption of equal slopes across categories of the independent (i.e., grouping) variable be satisfied. In all analyses of covariance performed in the present study, the slope of the covariate(s) in the independent variable categories are inspected for adherence to this statistical assumption. Results of the inspection are only reported if the assumption is violated.

transition training's overall efficacy in preparing them for line operations. Inspection of the Pearson product-moment correlations indicated that these three items are highly intercorrelated; the absolute value of the correlations for these three items exceeds .40 (Table 2 provides the correlation matrix for these and the other four items).

Although the three other evaluative items ("Recurrent training for my aircraft covers important abnormal procedures that are not encountered very often", "My company's computer-based training for this aircraft is effective", and "If [you were able to practice on a part-task FMS device], how useful did you find this type of training?") were also substantially correlated with the group of three items, they can be conceptually differentiated. The first item taps a specific area within the recurrent training domain, and the second and third items ask for a specific evaluation of training devices.

Table 2. Pearson Product-Moment Intercorrelation Matrix for Six Training Items.

| | Item 26 | Item 28 | Item 29 | Item 30 | Item 32 | Item 33 | Free-play |
|---|---|---|---|---|---|---|---|
| Item 26 | - - | .26** | .65** | -.42** | .23** | .40** | .21** |
| Item 28 | | - - | .26** | -.15** | .12** | .23** | .05* |
| Item 29 | | | - - | -.44** | .21** | .43** | .20** |
| Item 30 | | | | - - | -.12** | -.26** | -.14** |
| Item 32 | | | | | - - | .20** | .09** |
| Item 33 | | | | | | -- | .22** |
| Free-play | | | | | | | -- |

*Note.* Item 26 = "My transition training adequately prepared me for flying this aircraft on the line". Item 28 = "Recurrent training for my aircraft covers important abnormal procedures that are not encountered very often". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line". Item 32 = "During transition training, I was taught that the decision to use or not use automation on this aircraft is up to me". Item 33 = "My company's computer-based training for this aircraft is effective". The item "free-play" = "If [you were able to practice on a part-task FMS device], how useful did you find this type of training?"
*Note.* $^{*}p < .05$, $^{**}p < .001$.

Based on these conceptual differences, it was decided to analyze the group of three related items together in a MANCOVA, while the other evaluative items are analyzed separately in ANCOVAs. The three related training efficacy items are also utilized in analyses that seek to explicate the relationship between pilots' free-play training experience and their overall evaluation of training efficacy. The item indicating perceptions of automation use policy during training is also analyzed separately.

**Results: Training efficacy items MANCOVA**

Responses to the three overall training efficacy items were subjected to a three-factor (Organization X Fleet X Previous Automation Experience) MANCOVA, with flight experience utilized as a covariate. Flight experience was operationalized using total flying hours instead of total years in aviation because it is a 'purer' measure of actual flying experience – it is conceivable that two pilots, both with 20 years of aviation experience, could differ drastically in total flying hours.

Experience was a significant covariate, multivariate $F = 9.18$ (3,1502 *df*), $p < .001$. There was a significant and consistent relationship between total flying hours and responses to all three items. Experience was negatively correlated with responses to "My transition training adequately prepared me for flying this aircraft on the line" and "I gained an adequate understanding of the FMS during transition training" ($r = -.17$ and $-.15$, respectively; [1592 and 1579 *df*], $p < .001$), and positively correlated with "I learned most of what I need to know about this aircraft when I started flying it on the line" ($r = .13$ [1588 *df*]; $p < .001$). In other words, all else being equal, more experienced pilots evaluated their transition training more negatively than did their less-experienced (and presumably younger) counterparts (the relationship between age and evaluations of training is explored more fully immediately following the description of this MANCOVA).Both Airline and Fleet had a significant main effect on item responses; i.e., evaluations of training differed across both airlines and aircraft types. Responses across airlines differed significantly for all three items, multivariate $F = 3.24$ (18,4502 *df*), $p < .001$. Responses across fleets also differed significantly for the three items, multivariate $F = 6.20$ (18, 4502 *df*), $p < .001$. Additionally, the Airline X Fleet interaction was significant, $F = 2.69$ (21,4502 *df*), $p < .001$. The interaction between Airline and Fleet reflects the fact that the relative ratings for fleets *within* airlines differ when examined *across* airlines. None of the other interaction terms were significant. Table 3 provides the range of adjusted mean scores[11] on each item across airlines, across fleets within airlines, and across two fleets for airlines operating two widely used aircraft types, the B737-300 and the B757/767.

Table 3. Range of Mean Scores (overall SD) for Three Training Evaluation Items, for Seven Airlines, Across All Fleets and For Two Selected Fleets.

| | | Airline | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Item | Fleet | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Item 26 | All fleets | 3.22-3.50 (1.32) | --- | 2.91-3.46 (1.37) | 4.04-4.31 (1.10) | --- | 3.06-3.97 (1.30) | 3.21-3.93 (1.29) |
| | B73-300 | --- | 3.65 (1.31) | --- | --- | --- | 3.59 (1.12) | 3.21 (1.33) |
| | B75/767 | 3.22 (1.33) | --- | 3.46 (1.29) | 4.31 (1.10) | 3.65 (1.05) | 3.17 (1.29) | 3.93 (1.19) |
| Item 29 | All fleets | 3.06-3.16 (1.23) | -- | 3.05-3.18 (1.39) | 3.80-3.98 (1.18) | --- | 2.88-3.80 (1.32) | 2.66-3.64 (1.37) |
| | B73-300 | --- | 3.52 (1.18) | --- | --- | --- | 3.36 (1.28) | 2.66 (1.27) |
| | B75/767 | 3.16 (1.28) | --- | 3.05 (1.43) | 3.98 (1.21) | 3.47 (1.14) | 2.88 (1.53) | 3.64 (1.38) |
| Item 30 | All fleets | 3.54-3.68 (1.21) | -- | 3.60-3.95 (1.25) | 2.78-3.12 (1.37) | --- | 3.33-3.84 (1.29) | 3.20-4.27 (1.25) |
| | B73-300 | --- | 3.53 (1.24) | --- | --- | --- | 3.76 (1.28) | 4.27 (1.10) |
| | B75/767 | 3.68 (1.14) | --- | 3.95 (1.19) | 3.06 (1.45) | 3.43 (1.20) | 3.84 (1.20) | 4.20 (1.35) |

*Note.* Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".
*Note.* Airlines without agreement and disagreement ranges (Airlines 2 and 5) are represented in the sample by only one automated fleet.
*Note*. Means have been adjusted for the effect of the covariate. Unadjusted means are in Appendix B, Table B1.

To understand these results more fully, several strategies could be employed. In traditional discussions of ANOVA-type results, means and standard deviations are usually provided and discussed. This convention will be followed in the present study. However, to furnish additional descriptive information, for certain analyses the data

[11] For this and all other analyses where adjusted means are reported, unadjusted means are reported in Appendix B.

are described categorically. For the five-point Likert-scaled items, the five-option response scale will be collapsed into three categories reflecting agreement, neutrality, and disagreement, and the percent agreeing and/or percent disagreeing with an item across relevant subgroups will be provided.

Inspection of the categorical responses for the individual items by airline and fleet within airline (unadjusted for the effect of the covariate) underscores the fact that there is considerable variability across these factors. The response patterns also draw attention to the fact that (as observed by the FAA Human Factors Team, 1996), a non-trivial number of pilots report feeling that they were not well trained. Overall, 21% of pilots felt that they were not well prepared for line flying after transition training, and 26% felt that they did not gain an adequate understanding of the FMS during training. Table 4 presents the range of agreement and disagreement for the three relevant items across airlines, as well as agreement and disagreement across airlines for the B737-300 and B757/767 fleets.

Table 4. Range of Percent Agreeing (Disagreeing) with Three Training Evaluation Items for Seven Airlines, Across All Fleets and For Two Selected Fleets.

| | | Airline | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Item | Fleet | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Item 26 | All fleets | 42-66% (28-37%) | --- | 40-61% (26-51%) | 74-81% (10-15%) | --- | 40-80% (4-39%) | 51-70% (22-35%) |
| | B73-300 | --- | 69% (25%) | --- | --- | --- | 65% (21%) | 51% (35%) |
| | B75/767 | 42% (37%) | --- | 61% (26%) | 81% (10%) | 74% (17%) | 51% (39%) | 70% (22%) |
| Item 29 | All fleets | 42-45% (33-37%) | --- | 47-51% (36-41%) | 71-76% (16-19%) | --- | 42-79% (9-47%) | 37-59% (33-53%) |
| | B73-300 | --- | 66% (24%) | --- | --- | --- | 58% (28%) | 37% (53%) |
| | B75/767 | 45% (37%) | --- | 51% (41%) | 76% (18%) | 57% (26%) | 44% (47%) | 59% (33%) |
| Item 30 | All fleets | 54-61% (26-32%) | --- | 62-69% (19-27%) | 35-47% (42-51%) | --- | 35-70% (17-39%) | 57-74% (16-26%) |
| | B73-300 | --- | 55% (29%) | --- | --- | --- | 63% (25%) | 74% (16%) |
| | B75/767 | 61% (26%) | --- | 69% (19%) | 42% (43%) | 56% (29%) | 70% (17%) | 57% (26%) |

*Note.* Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".
*Note.* Airlines without agreement and disagreement ranges (Airlines 2 and 5) are represented in the sample by only one automated fleet.

To further understand the considerable differences in reported training efficacy, it helps to examine the responses for a specific aircraft type across several carriers. Inspection of the responses across the most common aircraft type, the B757/767, reveals that only 10% of this aircraft's pilots at Airline 4 felt that training did not adequately prepare them for line operations. In contrast, almost 40% of B757/767 pilots at Airline 6 felt that training did not adequately prepare them for the line. This is a strong indication that among different major U.S. airlines, flight training for the same aircraft may have radically different outcomes.

Similarly, 18% of B757/767 pilots at Airline 4 felt that they did not gain an understanding of the FMS; while 47% of all B757/767 pilots at Airline 6 felt this way. Finally, 70% of pilots from Airline 6's B757/767 fleet agreed that they learned most of what they needed to know about their aircraft when they actually started flying it; 40% of B757/767 pilots at Airline 4 felt this way. Slightly less striking differences in training evaluations can also be seen across the three carriers operating the B737-300.

Examining the responses across fleets *within* an airline also reveals much useful information. At Airline 6, only 40% of pilots in one fleet felt that transition training adequately prepared them for line flying, while in another fleet at this airline, this percentage was 80%. A similar range of agreement was observed for Airline 6 fleet responses to the item "I gained an adequate understanding of the FMS during transition training (i.e., 42% - 79%), and for responses to the item "I learned most of what I need to know about this aircraft when I started flying it on the line" (i.e., 35% - 70%). Smaller but noticeable differences in responses across fleets can also be observed within Airlines 1, 3, 4, and 7. This further underscores the supposition that, even within a single airline, some fleets provide better training than do others. This complements findings by Helmreich, Hines and Wilhelm (1996) detailing considerable differences in pilot performance across fleets within airlines.

As a final note, it can be observed that responses from Airlines 1 and 5 together did not seem to differ systematically from the other airlines (i.e., training evaluations from these two groups combined were not strikingly higher or lower than ratings of the other airlines combined). This indicates that there was no discernable systematic response bias due to whether pilots received the survey from their company or their union.

**Training efficacy items MANOVA across age groups.** As a further examination of the negative relationship between respondent experience and evaluations of transition training, respondents were grouped into four age cohorts using total years in aviation as a proxy for respondent age. The groups roughly corresponded to cohorts representing respondents of ages 20 to 30, 30 to 40, 40 to 50, and 50 to 60. This was accomplished by grouping those reporting from 0 to 10 years of aviation experience into a category representing the age 20 to 30 cohort, grouping those reporting more than 10 and up to 20 years of aviation experience into a category representing the age 30 to 40 cohort, etc. There were 226 respondents in the first group, 558 in the second group, 523 in the third group, and 363 in the fourth group[12].

Responses to the three overall training efficacy items were subjected to a single-factor (Age) MANOVA. Age had a significant main effect on responses to all three item; multivariate $F = 7.38$ (9,4898 *df*), $p < .001$. Inspection of the individual item mean scores by age grouping indicated a fairly unambiguous linear trend. For example, the mean response to item 29, "I gained an adequate understanding of the FMS during transition training", for the age 20 to 30 group was 3.86; for the 30 to 40 group, 3.67; for the 40 to 50 group, 3.43; and for the 50 to 60 group, 3.26. A similar pattern was observed in responses to the two other items. Table 5 provides the age group means.

Table 5. Mean Scores (SD) Across Age Groupings, Three Training Evaluation Items.

| Item | Age 20-30 | Age 30-40 | Age 40-50 | Age 50-60 |
|---|---|---|---|---|
| Item 26 | 4.18 (1.04) | 3.93 (1.19) | 3.58 (1.29) | 3.51 (1.37) |
| Item 29 | 3.85 (1.07) | 3.68 (1.19) | 3.43 (1.35) | 3.26 (1.38) |
| Item 30 | 3.14 (1.30) | 3.29 (1.30) | 3.54 (1.28) | 3.55 (1.29) |

*Note.* Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".

[12] This was done because the survey did not prompt for respondents' age, and most pilots enter the field of aviation at about age 20. Although this is only a rough approximation of the respondents' actual age, it is sufficient for the purposes of examining training evaluations across molar age categories.

Until now, this relationship between age and reactions to training has had much anecdotal but very little empirical support. This analysis suggests that there may be some truth to the commonly held stereotype concerning the supposed difficulty of older pilots in learning computer-related tasks[13]. However, other competing explanations do exist. For example, older, more experienced pilots may rate training more negatively because, on average, they have been through comparatively more transition training programs, and are more discriminating consumers of training. In short, it can be said that while there is an age effect, it cannot be definitively attributed to a particular phenomenon.

**Results: Other training evaluation item ANCOVAs**

**Recurrent training.** Responses to item 28, "Recurrent training for my aircraft covers important abnormal procedures that are not encountered very often" were subjected to a three-factor (Organization X Fleet X Previous Automation Experience) ANCOVA, with flight experience utilized as a covariate. There were no significant effects across any of the factors, or for the covariate.

**Computer-based training.** Responses to item 33, "My company's computer-based training for this aircraft is effective" were subjected to a three-factor (Organization X Fleet X Previous Automation Experience) ANCOVA, with flight experience utilized as a covariate. Both Airline and Fleet showed significant effects (for Airline, $F = 7.35$ [6,1495 *df*], $p < .001$; for Fleet; $F = 15.29$ [6,1495 *df*], $p < .001$). The Airline X Fleet interaction was significant; $F = 2.86$ (7,1495 *df*), $p < .01$. None of the other interaction terms or the covariate was significant.

The significant findings for aircraft and fleet were expected, because different fleets can (and do) utilize different computer-based training devices, and different airlines can choose from a number of different devices for the same fleet. The interaction between Airline and Fleet reflects the fact that the relative ratings for fleets *within* airlines differ when examined *across* airlines. For example, Airline 6 and Airline 7 both operate the B737-300 and the B757/767. At Airline 6, CBT is rated better by the B737-300 fleet than by the B757/767 fleet. However, at Airline 7, the reverse holds true. Table 6 presents the mean scores for this item across airlines and selected aircraft types.

[13] If the negative relationship between age and training evaluations is in fact due to this supposed difficulty, it is entirely possible that this relationship will disappear over time, as the last of the pre-personal computer revolution pilots retire and younger pilots take their places. The wisdom, judgement, and experience of this older generation will be sorely missed.

Table 6. Range of Mean Scores (overall SD) for the Item "My company's computer-based training for this aircraft is effective", Across All Fleets and for Two Selected Fleets.

| | Airline | | | | | | |
|---|---|---|---|---|---|---|---|
| Fleet | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| All fleets | 3.32-3.43 (1.17) | --- | 2.87-3.07 (1.31) | 3.36-3.58 (1.25) | --- | 2.37-3.88 (1.39) | 2.94-3.39 (1.32) |
| B73-300 | --- | 3.20 (1.16) | --- | --- | --- | 3.35 (1.26) | 2.94 (1.24) |
| B75/767 | 3.32 (1.28) | --- | 3.07 (1.31) | 3.36 (1.35) | 3.16 (1.21) | 2.37 (1.32) | 3.39 (1.36) |

*Note.* Item 26= "My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".
*Note.* Airlines without mean score ranges (Airlines 2 and 5) are represented in the sample by only one automated fleet.

**Free-play.** This section describes analyses regarding free-play, defined as the opportunity for non-jeopardy, individual, discretionary practice on various FMS training devices, during and in addition to regular transition training. Responses to the item "If [you were able to practice on a part-task FMS device], how useful did you find this type of training?" were subjected to a three-factor (Organization X Fleet X Previous Automation Experience) ANCOVA, with flight experience utilized as a covariate. It should be recalled that approximately 80% of respondents ($n$ = 1,413) reported having an opportunity for free-play on some type of part-task FMS device; therefore, this analysis only applies to this subset of the sample.

Experience was not a significant covariate in this analysis. Both Airline and Fleet had a significant effect on item responses; i.e., evaluations of free-play usefulness differed across both airlines and aircraft types (for Airline, $F$ = 5.90 [6,1288 $df$], $p < .001$; for Fleet, $F$ = 7.82 [6,1288 $df$], $p < .001$). The Airline X Fleet interaction was also significant, $F$ = 5.97 (7,1288 $df$), $p < .001$. These results reflect the fact that, for the same aircraft type, many airlines have different implementations of free-play. That is, many airlines designate different part-task FMS trainers as free-play devices, and the availability of these devices differs across airlines (i.e., some devices may be dedicated solely to free-play, while others may only be available for free-play after regular training hours). None of the other interaction terms were significant. Table 7 provides the range of mean scores across fleets for the airlines and for the selected fleets.

Table 7. Range of Mean Scores (SD) for the Item "If [you were able to practice on a part-task FMS device], how useful did you find this type of training?" for Seven Airlines, Across All Fleets and for Two Selected Fleets.

| | Airline | | | | | | |
|---|---|---|---|---|---|---|---|
| Fleet | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| All fleets | 3.19-3.30 (.76) | --- | 3.55-3.56 (.67) | 3.31-3.66 (.78) | --- | 2.91-3.65 (.83) | 3.25-3.58 (.83) |
| B737-300 | --- | 3.65 (.63) | --- | --- | --- | 3.48 (.78) | 3.25 (.81) |
| B75/767 | 3.30 (.78) | --- | 3.55 (.68) | 3.66 (.61) | 3.61 (.72) | 2.91 (.93) | 3.58 (.83) |

*Note.* Airlines without mean score ranges (i.e., Airlines 2 and 5) are represented in the sample by only one automated fleet.

Inspection of the mean scores on this item across carriers and aircraft types reveals very high endorsement of free-play overall (recall that this item is scaled from 1 to 4), although the variation in ratings indicates that some implementations of free-play are undoubtedly more effective than others. Still, this is encouraging news for training departments; this high endorsement of free-play suggests that increasing both the availability and quality of free-play may be an effective means of bolstering pilots' knowledge and skills with regard to automation.

Finally, an illustrative example of the Airline X Aircraft type interaction can be observed in the mean scores for the B737 and B757/767 fleets at Airlines 6 and 7. For Airline 6, the B737-300 fleet averages 3.48 on this item, while the B757/767 fleet averages 2.91. In Airline 7, the opposite pattern can be seen; the B737-300 fleet shows a lower average score than does the Airline 7 B757/767 fleet.

**Results: Training efficacy and free-play evaluation relationships**

The theoretical outlook described earlier suggests that those pilots who have the opportunity for free-play on a part-task FMS device will feel that they have a better understanding of the FMS, and leave training better prepared for line operations. Inspection of the intercorrelation matrix for the six training items presented in Table 2 provides some initial evidence that there is a relationship between ratings of free-play usefulness and evaluations of training efficacy. However, before attempting further analysis examining the relationship between free-play evaluations and training evaluations, it was decided to first establish whether there was a significant difference in training evaluations across groups reporting free-play use and those reporting no free-play use. Both of these analyses were necessary because providing a rating of free-play usefulness was contingent upon having responded "yes" to the item "Were you able to practice on a CDU/FMS part-task trainer". Therefore, the first analysis is intended to establish whether there was any discernible free-play effect, the second analysis is intended to establish whether there was a relationship between perceived free-play usefulness and training ratings.

To accomplish this, responses of free-play and no-free-play groups were compared on the three general training efficacy items. Before describing the analyses further, a statistical aside is necessary. Comparing training evaluations across those who used free-play and those who did not use free-play, and comparing training evaluations across categories of free-play usefulness, are conceptually distinct analyses. However, it was desirable to avoid computationally similar tests on the same data, as not doing so would increase the chance of committing a Type I error. Therefore, half of the respondents in the dataset were randomly selected, and utilized as a subsample for the first analysis. The unselected subsample was used in the analysis comparing training evaluations across free-play usefulness categories.

Responses to the three overall training efficacy items were subjected to a single-factor (Free-play utilization) MANCOVA, with flight experience as a covariate. After adjusting for the significant influence of the covariate, free-play utilization had a significant effect on item responses; multivariate $F = 7.95$ (3,761 *df*), $p < .001$. Inspection of the unadjusted means indicated support for the free-play hypothesis – those who reported being able to practice on some type of FMS part-task trainer also reported feeling better prepared for line flying, gaining a better understanding the FMS, and feeling that they learned more in training. Put another way, 71% of respondents who were able to practice on a part-task FMS device agreed that transition training prepared them for line operations. Only 56% of respondents who were not able to practice felt this way (the percent disagreeing with the item was 19% for those who practiced and 35% for those who didn't). Table 8 presents the adjusted mean scores across groups.

Table 8. Mean Scores (SD) for Free-play and No Free-play Groups, Three Training Evaluation Items.

| Item | Free-play | No free-play |
|---|---|---|
| Item 26 | 3.83 (1.21) | 3.32 (1.46) |
| Item 29 | 3.59 (1.25) | 3.01 (1.43) |
| Item 30 | 3.37 (1.28) | 3.77 (1.32) |

*Note.* Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".
*Note*. Means have been adjusted for the effect of the covariate. Unadjusted means are in Appendix B, Table B2.

Similarly, 'only' 23% of pilots who were able to practice felt that they did not gain an adequate understanding of the FMS, while among pilots who were not able to practice, the percentage who reported not gaining adequate FMS understanding was 41%.

To further investigate the relationship of free-play with training evaluations, responses of the remaining half of the sample to the three overall training efficacy items were subjected to a single-factor (Free-play usefulness) MANCOVA, with flight experience as a covariate. The results were also quite encouraging. After adjusting for the effect of experience, free-play usefulness had a significant effect on item responses, multivariate $F$ = 4.30 (9,2033 *df*), $p < .001$. Table 9 contains the adjusted means for the comparison groups.

Table 9. Mean Scores (SD) Across Free-play Usefulness, Three Training Evaluation Items.

| Item | Not at all useful | Slightly useful | Somewhat useful | Very useful |
|---|---|---|---|---|
| Item 26 | 2.86 (1.08) | 3.37 (1.42) | 3.67 (1.29) | 3.99 (1.14) |
| Item 29 | 2.49 (0.93) | 3.20 (1.41) | 3.52 (1.19) | 3.78 (1.19) |
| Item 30 | 3.86 (1.38) | 3.72 (1.29) | 3.48 (1.28) | 3.27 (1.28) |

*Note.* Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".
*Note*. Means have been adjusted for the effect of the covariate. Unadjusted means are in Appendix B, Table B3.

Inspection of the mean responses across free-play usefulness categories indicated increasingly more positive ratings accompanying higher ratings of free-play usefulness. Simply put, as pilots evaluated free-play more positively, their evaluations of transition training became more positive (i.e., they felt that they gained an adequate understanding of the FMS, and that training prepared them for line operations).

As a final illustration of this effect, it is instructive to once again examine the collapsed categorical responses to the training efficacy items. Only 42% of pilots who did not see free-play as useful felt that training adequately prepared them for line flying. This percentage rose to 57% among pilots who saw free-play as slightly useful, to 67% among pilots who saw free-play as somewhat useful, and to 77% among pilots was saw free-play as very useful. Similarly, only 25% of pilots who saw free-play as not useful felt they gained an adequate understanding of their aircraft's FMS; this percentage increased across categories to 49%, 58%, and 70%, respectively.

**Results: Perceptions of automation use policy during training ANCOVA**

Responses to the item "During transition training, I was taught that the decision to use or not use automation on this aircraft was up to me" were subjected to a three-factor (Organization X Fleet X Previous Automation Experience) ANCOVA, with flight experience utilized as a covariate. There was only one slightly significant effect, for Fleet (*F* = 2.46 [6,1520 *df*], *p* < .05). There were no other significant main effects or interactions, and the covariate was not significant. Post-hoc pairwise comparisons indicated that this significant result was due almost entirely to the low overall score given by B747-400 pilots. This fleet significantly differed from four other fleets; no other pairwise comparisons were significant. Overall, this was weak evidence for pervasive differences across fleets. Table 10 contains mean scores by aircraft type.

Table 10. Mean Scores (SD) for Responses to the Item "During transition training, I was taught that the decision to use or not use automation on this aircraft was up to me", Across Aircraft Types.

| | A320 | B737-300 | B747-400 | B75/767 | B777 | MD11 | MD80 |
|---|---|---|---|---|---|---|---|
| Item 32 | $3.07_{a}$ (1.43) | $3.29_{b}$ (1.34) | $2.49_{a,b,c,d}$ (1.39) | $3.05_{c}$ (1.41) | $3.14_{d}$ (1.30) | 3.25 (1.47) | 3.03 (1.35) |

*Note.* Means sharing the same subscript are significantly different from one another at p < .05, Dunn-Sidak pairwise test.

More interesting were the overall collapsed categorical responses to this item, as well as the correlations of this item with the general training efficacy items (presented earlier in Table 2). Inspection of the categorical responses revealed consensus across ANCOVA factors, but considerable individual variability. Across all respondents, 46% felt that they were taught to use automation in a discretionary manner and 43% felt that they were not. This indicates that air crews are leaving training with widely divergent views about discretionary automation use.

Furthermore, responses to this item were significantly correlated (all *p* < .001) with the three grouped measures of training efficacy and the three separate training evaluation items (see Table 2). This suggests that the perception of being trained for discretionary automation use was associated with higher ratings of training. Although this analysis does not allow assignment of a causal arrow, the phrasing of the items themselves suggests that attempting to develop better automation use judgment skills by teaching pilots to use automation in a discretionary manner may lead to better training outcomes. This implies that one way airlines could improve training outcomes is by ensuring that training endorses discretionary automation use; i.e., airlines should promulgate a discretionary automation use philosophy during training (Degani & Wiener, 1994).

**Special analyses: The effect of AQP curriculum change within fleets**

This section describes attempts to ascertain whether changes in a given fleet's training efficacy ratings occurred from pre- to post-transition to AQP-based standards. By contacting fleet managers at the airlines represented by the respondents, it was learned that two fleets represented in the sample had undergone transition from traditional to AQP standards; one was a B757/767 fleet, and the other was a B737-300 fleet[14]. Within each of the fleets groups were identified that had undergone AQP training and had undergone traditional training. For the B757/767 fleet, 12 respondents were AQP trained and 65 were not. For the B737-300 fleet, 76 were AQP trained and 59 were not.

To ascertain whether responses of the AQP and non-AQP groups differed within the respective fleets, responses to the three overall training efficacy items within each fleet were subjected to a single-factor (Training type) MANCOVA, with flight experience utilized as a covariate. For both comparisons, there was no significant difference in training ratings across AQP and non-AQP groups, but power to detect a difference was well below .80 for both comparisons. This calls into question the ability of these analyses to find a significant result if one truly

[14] All other fleets in the sample were either not AQP fleets, or had begun initial operation as AQP fleets (i.e., they had never been subject to traditional standards).

exists. In short, these analyses were equivocal; in the present study no conclusions can be offered regarding the efficacy of AQP-based training.

### Summary

The analyses presented in this section showed that overall, about one-quarter of pilots feel that transition training for their current automated aircraft did not adequately prepare them for the line, and did not provide an adequate understanding of the flight management system. However, ratings of training efficacy were shown to differ substantially across airlines, across airlines within single aircraft types, and within airlines across aircraft types. The analyses also showed that free-play, the use of a part-task FMS training device in a discretionary, non-jeopardy manner, was seen as highly useful by pilots, although there were differences in free-play ratings across airlines and fleets. Furthermore, those who utilized free-play reported better training outcomes than those who didn't, and higher free-play usefulness ratings were associated with higher ratings of training.

Additionally, it was shown that air crews leave training with widely divergent views about the appropriateness of discretionary automation use. Further analysis suggested that fostering judgment skills by training pilots in discretionary automation use might lead to better training outcomes. This points to the need for airlines to articulate and disseminate a philosophy of automation use that promotes discretionary use of automation.

A slight negative relationship was found between experience and ratings of training; generally, older, more experienced pilots evaluate transition training for their current automated aircraft slightly more negatively than do younger, less experienced pilots.

Finally, the relationship of Advanced Qualification Program-based curriculum implementation with training ratings was investigated, but the analyses were not sufficiently powerful to allow determination of whether the implementation of AQP standards in a training program was associated with improved ratings of training efficacy.

## METHOD AND RESULTS FOR EVALUATIONS OF EQUIPMENT

### Analytic method

**Independent variables, covariates, and correlates**. As in evaluations of training, multiple factors can influence pilots' evaluations of the automated equipment on their current aircraft. Specific factors likely to affect pilots' evaluations of equipment include their fleet, and their previous experience (if any) with automated aircraft. These variables are utilized as independent variables in the analyses described below.

Additionally, pilots' level of experience with their current aircraft, overall experience levels, and their evaluations of training might affect evaluations of automation. Similar to the analyses in the previous sections, these variables will be correlated with evaluations of automated equipment, and utilized as covariates where appropriate.

**Dependent variables**. The equipment evaluation items of the survey should be considered an item pool. Of the 28 items, many are quite similar (consider for example the items "The error messages I receive from the FMS are helpful" and "The error messages I receive from the FMS are useful in correcting a problem"). A number of other items are also conceptually related.

Because many items in this set were written to tap the same clearly defined concepts, it was decided to use factor analytic techniques to reduce the data. This allows the development of empirically sound composites for use in comparing across levels of the relevant independent variables. Both exploratory and confirmatory techniques were employed in the present study (exploratory factor analysis was performed on a randomly selected half of the dataset, confirmatory on the remaining half).

Not only are many items related, most of the item *categories* are somewhat related. Therefore, it could be expected that item responses in many of these areas would also be related. For this reason, principle-axis factor analysis was first attempted with no prior expectation for the number of and structure of the factors. Subsequent solutions attempted to converge on a viable factor structure; finally, confirmatory factor analysis was used to corroborate the factors derived using exploratory techniques (Joreskog & Sorbom, 1989).

It should be noted that, while the use of item composites is certainly useful in comparative analyses, many of the individual items from this item pool pertain to critical areas of concern identified by ALPA (1996), the ATA (in press), and the FAA (1996). These areas include the ease with which needed information can be retrieved from the aircraft's FMS, the diagnosticity and usability of FMS-generated errors and warnings, the degree to which information in the FMS is logically presented, and the ease with which the aircraft can be transitioned from automated to manual flight. Because the information generated from analysis of items covering these topics should prove very useful, both scale-level and item-level analyses are provided. As in the previous section, analysis of covariance techniques will be utilized where appropriate.

**Results: Equipment item scale development**

Using a randomly selected half of the sample, the correlation matrix of 28 equipment evaluation items was subjected to unconstrained principle-axis factor (PAF) analysis, with oblique rotation of the solution. The analysis yielded a three-factor solution. After rotation, the three-factor solution yielded two factors with eigenvalues > 1.0 (these two factors accounted for 32% of the variance). The eigenvalue for the third factor was 0.74. Fourteen items loaded significantly onto the first factor; three onto the second factor. The third factor had only one significant item loading. This factor was removed from further consideration. For the first factor, Cronbach alpha for the ten items contributing positively to reliability was .84. The four items with the lowest loadings negatively affected the reliability coefficient of this factor, so they were eliminated from further consideration. Reliability for the three items loading onto the second factor was .80, with all three items contributing to reliability. The solution is detailed in Table 11 below.

Inspection of the items comprising the first factor revealed a common theme; all ten items probed pilots' evaluations of the automation's ease of use, its ability to provide useful, diagnostic information, and the degree to which it fostered flight crew situation awareness. In short, the items tapped a global construct indicating the degree to which a particular aircraft's automated systems helped or hindered crews perform their flight tasks. The factor was named "Automation-Crew Synergy". Scores on this unit-weighted scale can range between 10 and 50. The second factor consisted of three items tapping the degree to which the error messages and warnings generated by the aircraft's FMS were helpful in resolving problems during flight. This factor was named "Troubleshooting and Problem Solving". Scores on this unit-weighted scale can range between 3 and 15.

Table 11. Results of Factor Analyses, Equipment Items.

| Item | *F* 1 | *F* 2 | *F* 3 | Λ | Θ |
|---|---|---|---|---|---|
| 4. Always apparent when an automated system fails | **.68** | | .28 | **.55** | **.70** |
| 15. Always apparent when a system acts other than expected | **.60** | | | **.55** | **.69** |
| 6. This FMS provides me w/ just the right amount of info | **.59** | | -.17 | **.74** | **.46** |
| 12. If system fails, I understand nature of the failure quickly | **.57** | -.12 | | **.60** | **.65** |
| 13. This aircraft was designed to keep pilot ‘in the loop’ | **.52** | | -.26 | **.66** | **.57** |
| 7. FMS programming features are well designed | **.52** | | -.31 | **.71** | **.50** |
| 2. Satisfied with the format of the displays on this aircraft | **.51** | | -.16 | **.70** | **.51** |
| 1. This FMS is easy to use | **.51** | | -.40 | **.71** | **.49** |
| 8. The autothrottle system executes changes accurately | **.51** | | | **.43** | **.82** |
| 22. FMS information is easily accessible | **.50** | -.17 | -.27 | **.76** | **.43** |
| 5. FMS error messages are easy to understand | .47 | -.33 | .12 | | |
| 3. The autothrottle system executes changes smoothly | .46 | | | | |
| 11. I’m always aware of changes commanded by autothrottles | .46 | | -.15 | | |
| 19. I find all modes of the FMS useful | .45 | -.16 | -.11 | | |
| 9. It is very difficult to get needed info from this FMS | -.37 | .17 | .29 | | |
| 21. I try to use all the modes and features of this FMS | .36 | | | | |
| 10. Difficult to determine why inputs not accepted by FMS | -.36 | .18 | | | |
| 14. I get insufficient mode change feedback from automation | -.29 | | | | |
| 25. FMS error messages not useful in correcting a problem | | **-.86** | | **.87** | **.24** |
| 20. FMS error messages do not help me to solve a problem | | **.71** | | **-.76** | **.42** |
| 16. FMS error messages are helpful | | **-.70** | | **.70** | **.51** |
| 23. I receive useful thrust information from throttle position | | | -.46 | | |
| 24. FMS information is logically presented | .37 | -.26 | -.39 | | |
| 17. Difficult to transition from automated to manual flight | -.17 | | .32 | | |
| 27. This aircraft doesn’t make use of basic aviation skills | -.25 | | .26 | | |
| 34. I rely on my basic aviation skills when flying this aircraft | | | -.23 | | |
| 31. My past experience prepared for operating this aircraft | | | -.19 | | |
| 18. Expect immediate thrust change when I move the throttles | | | -.15 | | |

*Note.* *F* = factor loading for PAF with oblique factor rotation; Λ = CFA lambda; Θ = CFA theta. Items in CFA are in boldface. For PAF, items with loadings < |.10| are not shown.
*Note.* CFA for two target factors yielded a GFI = .93; $\chi^2$ = 364.26 (64 *df*), $p < .001$; $\chi^2/df$ = 5.69.

To cross-validate the factors in the remaining half of the sample, maximum-likelihood confirmatory factor analysis (with freely estimated factor correlation) specifying the two target factors was performed using the variance-covariance matrix of item responses. Inspection of the $\chi^2/df$ value (Byrne, 1989; Marsh & Hocever, 1985), and goodness-of-fit index indicated that the imposed factor structure was a good fit to the data (GFI = .93; $\chi^2$ = 364.26 (64 *df*), $p < .001$; $\chi^2/df$ = 5.69; lambda and theta values are also in Table 11). These two correlated factors ($r = .36$ [1647 *df*], $p < .001$) were retained and used in comparative analyses.

**Results: Correlations between experience, training, and equipment ratings**

The two scales derived above were correlated with measures of pilots’ total flight hours, flight hours in their current aircraft, and the three overall training evaluation items (items 26, 29, and 30). There were large and significant correlations between Automation-Crew Synergy and all three training evaluation items (correlations were .46, .46, and -.21, respectively; [1646, 1646, and 1643 *df*], $p < .001$). The Troubleshooting and Problem Solving scale was significantly correlated with items 26 and 29 (.16 and .18, respectively; [1656 *df*], $p < .001$). Correlations between the two scales and the demographic variables were below .10 in absolute value.

These training evaluation-equipment evaluation correlations are of considerable importance. They suggest that pilots' evaluations of training are strongly related to evaluations of their aircraft's automation. Assume for the moment that the causal arrow points from training evaluations toward equipment evaluations (a not unreasonable assumption, since training for use of the automation occurs before pilots can begin to develop informed judgements about the equipment). This assumption leads to the conclusion that better preparation for line operations and FMS use during training leads to better understanding of and greater satisfaction with the FMS, as well as higher evaluations of FMS usability, when pilots fly the line. In short, flight deck automation cannot be evaluated in tabula rasa fashion – it must be evaluated while taking into account the operator's training background. Finally, this finding also suggests that any cross-aircraft analysis of automation evaluations must control for the effect of training evaluations.

To explore these assertions in more detail (and to control for the effect of collinearity between the training items) two multiple regression analyses were performed, each attempting to predict scores on one equipment scale from ratings of training. The regression procedures utilized a forward stepwise selection procedure. Criteria for inclusion in the equation was set at $p < .05$; exclusion was set at $p < .10$. Table 12 shows that overall, training ratings significantly predict equipment evaluations. For the first analysis, training evaluations accounted for 25% of the variance in Automation-Crew Synergy scores ($F = 277.68$ [2,1642 *df*], $p < .001$). Standardized regression (beta) weights for the two items included in the equation (items 26 and 29) were .28 and .29, respectively. The multiple correlation (R) of these training evaluations with scores on Automation-Crew Synergy was .50.

Table 12. Summary of Regression Models Predicting Equipment Scale Scores From Training Ratings.

Model 1: Predicting Automation-Crew Synergy scale scores

| Training Items | Std. regression weight | |
|---|---|---|
| Item 26 | .28** | $F = 277.68$ (2,1642 *df*), $p < .001$ |
| Item 29 | .29** | R = .50 |
| Item 30 | .04 | $R^2 = .25$ |

Model 2: Predicting Troubleshooting and Problem Solving scale scores

| Training Items | Std. regression weight | |
|---|---|---|
| Item 26 | .09* | $F = 21.85$ (3,1651 *df*), $p < .001$ |
| Item 29 | .15** | R = .20 |
| Item 30 | .07* | $R^2 = .04$ |

*Note*. **p* < .05, ***p* < .001.

For the second analysis, training evaluations accounted for 4% of the variance in Troubleshooting and Problem Solving scores ($F = 21.85$ [3,1651 *df*], $p < .001$). Beta weights for the three items in the equation were .09 and .15, and .07, respectively. The multiple correlation of these training evaluations with scores on Troubleshooting and Problem Solving was .20. In general, the first regression analysis supports the assertion that pilots' training experiences exerts a fairly large effect on how they evaluate the automation on their aircraft, although the second analysis offers only weak supporting evidence for this.

**Results: Scale comparisons across aircraft types**

Responses to the two equipment evaluation scales were subjected to a two-factor (Fleet X Previous Automation Experience) MANCOVA, with training efficacy ratings utilized as covariates[15].

As expected, the three training evaluations were significant covariates, multivariate $F$ = 89.98 (6,3176 *df*), $p$ < .001. Fleet had a significant main effect on responses to the equipment evaluation scales, multivariate $F$ = 4.94 (12,3176 *df*), $p$ < .001. There was no significant effect for previous automation experience, or for the Fleet X Previous experience interaction. Table 13 contains the adjusted means by fleet.

Table 13. Means (SDs) for Two Equipment Evaluation Scales, Seven Aircraft.

| Aircraft | Automation-Crew Synergy | Troubleshooting & Problem Solving |
|---|---|---|
| A320 | 34.70<br>(7.80) | $9.67_{a}$<br>(1.29) |
| B737-300 | $34.49_{e}$<br>(7.70) | $9.27_{a,b,c}$<br>(1.49) |
| B747-400 | $36.42_{a,b}$<br>(8.61) | $9.82_{b}$<br>(1.46) |
| B757/767 | $35.98_{c,d,e}$<br>(7.40) | $9.71_{c}$<br>(1.43) |
| B777 | 35.01<br>(7.43) | 9.71<br>(1.42) |
| MD11 | $31.86_{a,c}$<br>(9.12) | 9.77<br>(1.42) |
| MD80 | $32.10_{b,d}$<br>(6.34) | 9.17<br>(1.36) |

*Note.* Scores sharing the same subscript differ significantly from one another at $p$ < .05, Dunn-Sidak pairwise test.
*Note*. Means are adjusted for the effects of the covariates. Unadjusted means are presented in Appendix B, Table B4.

Inspection of the mean scores for the scales (adjusted for the covariates), as well as the post-hoc tests presents an interesting picture. Boeing aircraft generally earn higher scores than McDonnell-Douglas aircraft on the Automation-Crew Synergy scale (the Airbus A320's scores are not significantly different from scores of Boeing aircraft on this scale). However, this pattern of scores is not evident in scores for Troubleshooting and Problem-Solving. Instead, there seems to be a generation effect, with earlier vintage aircraft (e.g., the Boeing 737-300) earning lower scores on this scale than did more advanced aircraft such as the A320, B747-400, B757/767 and B777. Although the early-vintage MD80 fleet also showed quite low scores, there were no *statistically* significant pairwise comparisons with this fleet and others according to the results of the Dunn-Sidak pairwise tests. This is most likely due to its larger standard error of difference value. These results are not surprising, as the newer aircraft from all manufacturers incorporate comprehensive alerting and warning systems that are considerable improvements over the fault alerting, diagnosis and management systems found on older automated aircraft.

[15] Airline was not utilized as a factor in comparisons because there was no reason to expect that equipment evaluations would differ across airlines (that is, once the effect of training was removed). As a check on this assumption, I compared the two equipment scale scores separately across carriers, for the four aircraft types operated by multiple airlines. The aircraft were the A320, B737-300, B747-400, and the B757/767. Thus, eight ANCOVAs were performed, covarying on training ratings. In support of the above assumption, half of the comparisons were not significant. Of the significant comparisons, only one exceeded $p$ < .001 (for scores on Troubleshooting and Problem Solving across two airlines flying the A320). The rest were only significant at $p$ < .05, which is weak evidence for an effect, considering the large group sizes.

Overall, these results suggest that both the automation generation hypothesis and the manufacturer hypothesis may be tenable. However, while the item composites are reliable indicators of flight crew overall evaluations, they do not lend themselves to detailed explorations of pilots' evaluations in more specific domains. To further investigate this, a series of ANCOVAs was performed which compared responses to selected equipment evaluation items across aircraft types, while controlling for the effects of training ratings.

**Results: Item comparisons across aircraft types**

**Item selection.** As mentioned earlier, the equipment evaluation items cover a number of critical areas identified by various industry and regulatory groups. Of the 28 items, six seemed to best represent some of the areas of concern. Items 14 ("I get insufficient feedback regarding mode changes from the automated systems") and 11 ("I am always aware of changes to thrust level that are commanded by the autothrottle system") indicate the degree to which the respondent's aircraft adequately informs the crew of autonomous, automation-induced changes to flight parameters. Items 9 ("It is very difficult to get the information I need from the FMS in this aircraft") and 24 ("Information I may need from the FMS is logically presented") serve as indicators of the 'user-friendliness' of an aircraft's automation. Item 25 ("The error messages I receive from the FMS are useful in correcting a problem") indicates the utility of a particular aircraft's error and warning messages. Finally, item 17 ("It is difficult to transition from automated to manual flight in this aircraft") is an indicator of the ease with which a particular aircraft can be transitioned from automated to manual flight.

**Item analyses.** Responses to the six equipment evaluation items were analyzed in a series of six single-factor (Fleet) ANCOVAs, with training efficacy items utilized as covariates. The results of the six analyses are displayed in Tables 14 and 15.

Table 14. Source Tables for Six Single-Factor ANCOVAs (equipment evaluation items), with Training Evaluation Covariates.

| Item | Source | SS | *df* | MS | *F* | *p* |
|---|---|---|---|---|---|---|
| 14. I get insufficient feedback regarding mode changes from the automated systems | Regression | 96.09 | 3 | 32.03 | 23.26 | .001 |
| | Fleet | 10.30 | 6 | 1.72 | 1.25 | --- |
| | Error | 2227.96 | 1618 | 1.38 | | |
| | Total | 2335.32 | 1627 | 1.44 | | |
| 11. I am always aware of changes to thrust level that are commanded by the autothrottle system. | Regression | 145.89 | 3 | 48.63 | 32.10 | .001 |
| | Fleet | 60.29 | 6 | 10.05 | 6.63 | .001 |
| | Error | 2451.33 | 1618 | 1.52 | | |
| | Total | 2644.24 | 1627 | 1.63 | | |
| 9.It is very difficult to get the information I need from the FMS in this aircraft. | Regression | 152.52 | 3 | 50.84 | 62.99 | .001 |
| | Fleet | 68.72 | 6 | 11.45 | 14.19 | .001 |
| | Error | 1305.14 | 1617 | 0.81 | | |
| | Total | 1538.17 | 1626 | 0.95 | | |
| 24. Information I may need from the FMS is logically presented | Regression | 276.25 | 3 | 92.08 | 101.89 | .001 |
| | Fleet | 65.17 | 6 | 10.86 | 12.02 | .001 |
| | Error | 1461.38 | 1617 | 0.90 | | |
| | Total | 1822.65 | 1626 | 1.12 | | |
| 25. The error messages I receive from the FMS are useful in correcting a problem | Regression | 151.93 | 3 | 50.64 | 51.31 | .001 |
| | Fleet | 42.12 | 6 | 7.02 | 7.11 | .001 |
| | Error | 1592.99 | 1614 | 0.99 | | |
| | Total | 1808.90 | 1623 | 1.11 | | |
| 17. It is difficult to transition from automated to manual flight in this aircraft | Regression | 105.07 | 3 | 35.02 | 29.80 | .001 |
| | Fleet | 53.74 | 6 | 8.96 | 7.62 | .001 |
| | Error | 1901.34 | 1618 | 1.18 | | |
| | Total | 2052.97 | 1627 | 1.26 | | |

*Note*. Regression = Regression = effect for covariate (i.e., training efficacy).

Table 15. Means (SDs) for Six Equipment Evaluation Items, Seven Aircraft Types.

| | Item | | | | | |
|---|---|---|---|---|---|---|
| Aircraft | Item 14 | Item 11 | Item 9 | Item 24 | Item 25 | Item 17 |
| A320 | 2.69 (1.24) | $3.17_{a,b,c}$ (1.34) | $1.94_{a,b,c}$ (.97) | $3.60_{a,b}$ (1.09) | $3.63_{a,b}$ (1.02) | $2.03_{a,b}$ (1.24) |
| B737-300 | 2.75 (1.18) | $3.60_{a}$ (1.24) | $1.76_{d,e}$ (.91) | $3.78_{c,d}$ (1.01) | $3.31_{a,c,d}$ (1.03) | $1.59_{a,c}$ (.95) |
| B747-400 | 2.70 (1.22) | $3.65_{b}$ (1.29) | $1.90_{f,g}$ (1.13) | $3.82_{e,f}$ (1.12) | $3.65_{c,e}$ (1.17) | 1.80 (1.13) |
| B757/767 | 2.68 (1.17) | $3.57_{c}$ (1.20) | $1.73_{a,h,i}$ (.86) | $3.86_{a,g,h}$ (.98) | $3.68_{d,f}$ (1.02) | $1.71_{b,d}$ (1.06) |
| B777 | 2.91 (1.32) | 3.47 (1.24) | $1.94_{j,k}$ (.88) | $3.72_{i}$ (.90) | 3.54 (.97) | 1.86 (1.05) |
| MD11 | 2.53 (1.09) | 3.18 (1.30) | $2.59_{b,d,f,h,j}$ (1.16) | $2.95_{b,c,e,g,i}$ (1.25) | 3.37 (1.20) | $2.29_{c,d}$ (1.17) |
| MD80 | 2.96 (1.07) | 3.28 (1.24) | $2.50_{c,e,g,i,k}$ (1.13) | $3.27_{d,f,h}$ (1.03) | $3.15_{b,e,f}$ (.95) | 1.77 (1.33) |

*Note.* Item 14 = "I get insufficient feedback regarding mode changes from the automated systems". Item 11 = "I am always aware of changes to thrust level that are commanded by the autothrottle system". Item 9 = "It is very difficult to get the information I need from the FMS in this aircraft". Item 24 = "Information I may need from the FMS is logically presented". Item 25 = "The error messages I receive from the FMS are useful in correcting a problem". Item 17 = "It is difficult to transition from automated to manual flight in this aircraft".
*Note.* Scores sharing the same subscript (for the same item) differ significantly from one another at $p < .05$, Dunn-Sidak pairwise test; means are adjusted for the effects of the covariates (unadjusted means are in Appendix B, Table B5).

As expected, training evaluations were significant covariates in all six analyses. There were no significant differences across aircraft types for item 14, "I get insufficient feedback regarding mode changes from the automated systems". There was a significant difference across aircraft types for item 11, "I am always aware of changes to thrust level that are commanded by the autothrottle system". Inspection of the adjusted means in Table 15 reveals that generally speaking, Boeing aircraft (the B737-300, B747-400, B757/767 and B777) earn the highest scores on this item; the Airbus A320, MD11, and MD80-series aircraft earn relatively lower scores (the Dunn-Sidak pairwise test, however, only identified the A320 and not the two McDonnell-Douglas aircraft as differing significantly from the Boeing types).

The lower thrust awareness rating for the A320 is not surprising, as this aircraft's thrust management system incorporates a different control interface than the other aircraft. Unlike the other aircraft represented in this study, the throttle levers on the A320 do not move in response to pilot- or FMS-initiated thrust changes. Many interested parties (e.g., pilots, their representative groups, and researchers) feel that this is not a synergistic design, as pilots have traditionally depended upon throttle movement as a cue for energy awareness when using autothrust systems. Interestingly, the most virulent criticism of the Airbus autothrust system has often come from those who do not operate Airbus aircraft. This finding represents some of the first evidence showing A320 pilots *themselves* feel that awareness of thrust management is lower in their aircraft than it is in other aircraft.

However, this finding may not be *necessarily* due to the fact that the A320 throttles do not move. Inspection of the mean scores for this item reveals that both the McDonnell-Douglas MD11 and MD80-series aircraft also earn relatively low scores on this item. Both of these aircraft incorporate back-driven (i.e., moving) autothrottles. If the fact that Airbus A320 throttles do not move were responsible for the low thrust awareness ratings found here, then it could be expected that the McDonnell-Douglas aircraft would receive higher ratings. This does not seem to be the case.

Complicating the issue further, the MD80-series aircraft represents one of the earliest autothrust system designs (circa early 1980's), while the MD11's systems represent some of the latest technological advances (this aircraft was put into service in the early 1990's). Thus, the differences in thrust awareness ratings are probably not due to a generation effect. Although the sample utilized in the present study is somewhat unbalanced (a common problem in field research of this type), it appears that the differences in responses to this item may be due to differences in how each manufacturer implements autothrust system annunciations and cues on their aircraft. While it is not possible to ascertain the exact nature of these differences in the present study, these results may spur future efforts to more fully determine the antecedents to enhanced or degraded thrust awareness on different aircraft.

The effect for aircraft type was significant for item 9, "It is very difficult to get the information I may need from the FMS in this aircraft". Inspection of the mean scores for this reverse-scored item (i.e., lower scores imply that it is not difficult to get needed information from the FMS) reveals that overall, although pilots in all aircraft types generally disagree with the item, some disagree more than others do. In particular, the two McDonnell-Douglas aircraft (the MD11 and MD80-series) earned the least favorable scores on this item. There was also a significant difference across aircraft type for responses to item 24, "Information I may need in the FMS is logically presented". Again, the MD11 and MD80-series earned relatively unfavorable scores on this item (additionally, the A320 is significantly lower than the highly rated B757/767 in this analysis). Taken together, the results of these two analyses suggest that there may be some commonality in McDonnell-Douglas automation implementations (even though these aircraft have different flight management systems) that causes pilots to more negatively evaluate the flight management systems of these aircraft.

Responses to item 25, "The error messages I receive from the FMS are useful in correcting a problem" differed across aircraft types. The A320 and newer Boeing aircraft (i.e., the B747-400, B757/767, and B777) earned higher scores than did the B737-300 and MD80-series. Although the two oldest aircraft (the MD80-series and the B737-300) earned the lowest scores on this item, suggesting a generation effect, scores for the MD11 were not much higher than scores for these two aircraft.

Finally, scores across aircraft type were significantly different for the reverse-scored item "It is difficult to transition from automated to manual flight in this aircraft." *Lower* scores (i.e., disagreement) for this item imply that it is relatively *easier* to transition from automated to manual flight. Ease of transition from automated to manual flight is important because the dynamic nature of flight often requires that air crews fly using many different combinations of automatic and manual control, especially during approaches. Inspection of the mean scores across aircraft type initially suggests a generation effect. Generally, pilots felt that it was harder to transition from automated to manual flight in the newer aircraft (in particular, the A320 and MD11), especially relative to older aircraft such as the B737-300. However, the newest aircraft in the sample, the B777, is seen as relatively easy to transition (it did not differ significantly from the older aircraft such as the B737-300), which casts doubt on the ascription to automation generation.

It should be noted that while these comparative analyses provide useful information regarding the relative strengths and weaknesses of various aircraft, an examination of the categorical responses (i.e., the absolute response levels) reveals that no one manufacturer has all the answers. For example, inspection of the categorical responses to item 14, an item with no significant differences in responses across aircraft types, indicates that between 22% and 34% of pilots feel that they in fact get insufficient mode change feedback from their aircraft's systems.

Although Boeing aircraft generally come out on top in these comparative analyses, categorical response inspections reveal that even the top-rated aircraft for a particular item still show absolute ratings that can be described as less than satisfactory. For example, 24% of B747-400 pilots do not feel that their aircraft provides them with sufficient feedback regarding thrust level changes commanded by the autothrottles. Similarly, 13% of B757/767 pilots do not feel that FMS information on their aircraft is logically presented.

## Summary

In this section two highly reliable, correlated factors were derived. One factor indicates the degree to which a particular aircraft's automated systems helped or hindered crews perform their flight tasks, the other measures the degree to which error messages and warnings generated by the aircraft's FMS were helpful in resolving problems during flight. The comparative analyses presented here suggest that in order to understand pilots' evaluations of the automation on their current aircraft, one must first take into account their training background.

In general, the automation on Boeing aircraft is seen as more conducive to effective crew-automation functioning than that found on Airbus A320 or McDonnell-Douglas MD11 and MD80-series aircraft, although there are several exceptions to this general conclusion. For example, both the Airbus A320 and the Boeing aircraft in the sample earn similar high scores for ease of gathering information from the FMS. Those flying the MD11 rate this aircraft high for providing mode change feedback. Also, the Airbus A320 and newer Boeings are seen as better at alerting and advising crews of systems problems, in contrast to the B737-300 and McDonnell-Douglas aircraft.

It was demonstrated that pilots' reported awareness of thrust management was slightly lower in Airbus and McDonnell-Douglas aircraft, although despite polemic to the contrary, the Airbus's non-moving autothrottles may not be the primary reason for this. Finally, Boeings and the MD80-series aircraft earn high ratings for ease of transitioning from automated to manual flight.

Some of the analyses also demonstrated generation effects, both positive and negative. In general, newer aircraft were seen as much more conducive to effective fault alerting, diagnosis, and management. However, some of these same aircraft (e.g., the A320 and MD11) were also perceived to be more difficult to transition from automated to manual flight control.

## METHOD AND RESULTS FOR AUTOMATION MANAGEMENT ITEMS

### Analytic method

**Independent variables, covariates, and correlates**. As outlined earlier, pilots' automation management attitudes may be related to flying experience, the need to avoid uncertain, ambiguous situations, and perceptions of company philosophy and policy of automation use. Automation management attitudes may also differ across organizations, fleets, and crew position. These variables are utilized as independent variables, covariates, and/or correlates where appropriate.

**Dependent variables**. Sixteen of the 20 items[16] discussed here are intended to tap a priori conceptual domains that derive from the earlier-described consequences of automation use. Several items measure pilots' endorsement of discretionary, flexible use of automation (e.g., "Automated systems should be used at the crews' discretion" and "I try to use automation as much as possible during flight operations", respectively, are positive and negative indicators of this domain). Others reflect the importance attached to communicating automation actions and intentions to the other crew member (e.g., "Automated cockpits require more verbal communication between crewmembers", and "I make sure the other pilot acknowledges programming changes I make in the FMC"). Additional items were written to indicate the level of concern with potential problems inherent to accomplishing a process via little-used automation functions, and with losing flying skills when using automation (e.g., "It's easy to forget how to do FMC operations that are not performed often", and "I am concerned that the use of automation will cause me to lose flying skills").

These items form an item pool; multiple items were intended to be indicators of the same construct. Therefore, factor analyses (both exploratory and confirmatory) were used to reduce the data to meaningful constructs indicative of the above domains. The techniques used to derive the latent constructs followed the same general methodological outline described previously. As in previous sections, examination of individual item responses is likely to be informative; therefore, both scale-level and item-level analyses will be performed.

[16] Recall that three items prompt for pilots' general preference for automation, and one item measures pilots' perceptions of company policy regarding automation use.

Sherman, Helmreich, and Merritt (in press) administered 16 of these 20 items to pilots in 12 nations (including several organizations within the U.S.) in an earlier study. This study found attitudes suggesting potential vulnerabilities in automation management among a considerable proportion of pilots in all nations. This work also found large differences in item endorsement across nations and relatively small differences across organizations within nations. These results have two implications: 1) the absolute level of item endorsement in a given sample may be an important indicator of potential safety threats; and 2) there may not be statistically significant differences in item responses across organizations within a given nation, although there is likely to be considerable variability at the individual level. Taken together, these points imply that even if no statistically significant differences are found across relevant independent variables, investigation of the overall item responses may serve to identify potential threats to safety. Therefore, the first portion of this section will be devoted to an examination of the overall item endorsement in this sample. Categorical responses to the items, as well as more detailed examination of selected items seen as particularly indicative of good automation management by Sherman, Helmreich, and Merritt, are described below.

**Results: Exploration of categorical responses**

Descriptive statistics (frequency distributions, means, and standard deviations) were calculated for the twenty automation management items. The collapsed categorical responses to the items, as well as item means and standard deviations, are presented in Table 16.

Table 16. Means, SDs and Collapsed Categorical Responses for Twenty Automation Management Items.

| Item | Mean | SD | Percent responding in each category: Disagree | Neutral | Agree |
|---|---|---|---|---|---|
| 35. Prefer flying automated aircraft | 4.16 | 1.01 | 8 | 13 | 79 |
| 36. I can rapidly access FMC info in abnormal conditions | 3.67 | 1.10 | 20 | 13 | 67 |
| 37. Effective crewmember always uses the automation | 3.47 | 1.23 | 26 | 16 | 58 |
| 38. Better to avoid FMC reprogramming in high workload | 3.72 | 1.13 | 19 | 15 | 66 |
| 39. Concerned that automation will cause loss of flying skills | 3.36 | 1.29 | 28 | 15 | 57 |
| 40. Easy to forget how to do FMC ops that are not done often | 3.73 | 1.07 | 16 | 15 | 69 |
| 41. I look forward to more automation | 2.85 | 1.21 | 39 | 32 | 29 |
| 42. Pilots should avoid disengaging automation | 1.92 | 1.11 | 76 | 11 | 13 |
| 43. There are FMC modes and features I don't understand | 2.86 | 1.32 | 44 | 14 | 42 |
| 44. Automated cockpits require more verbal communication | 3.92 | 1.04 | 12 | 17 | 71 |
| 45. I maintain flying proficiency by disengaging automation | 4.15 | 1.02 | 10 | 10 | 80 |
| 46. Automation leads to safer operations | 3.76 | 1.07 | 13 | 19 | 68 |
| 47. Automated cockpits require more cross-checking | 4.08 | 0.98 | 9 | 13 | 78 |
| 48. My company expects me to always use automation | 3.29 | 1.26 | 31 | 16 | 53 |
| 49. I make sure other pilot acknowledges my FMC changes | 4.24 | 0.94 | 8 | 9 | 83 |
| 50. I feel free to select level of automation at any given time | 4.36 | 0.93 | 7 | 6 | 87 |
| 51. Automation should be used at crew's discretion | 3.94 | 1.18 | 18 | 11 | 71 |
| 52. Flying automated aircraft alters transfer of information | 4.13 | 0.91 | 7 | 12 | 81 |
| 53. I try to use automation as much as possible | 3.65 | 1.11 | 18 | 16 | 66 |
| 54. Difficult to know what FMC ops others are performing | 2.18 | 1.12 | 71 | 12 | 17 |

Inspection of the categorical responses provides an interesting picture of air crew attitudes toward automation management. Overall, it appears that a non-trivial proportion of pilots do not endorse what can be described as safe automation management practices. For example, Sherman, Helmreich, and Merritt (in press) reported that a significant minority of all U.S. pilots (i.e., roughly 33%) did *not* endorse avoidance of reprogramming the FMS as a means of avoiding high workload (item 38). The present study shows that while a majority (66%) of pilots agreed that it is better to avoid reprogramming the FMS when workload increases, a considerable number of pilots (19%) do not agree with this item.

Sherman, Helmreich, and Merritt also reported that a majority (about 70%) of U.S. pilots endorsed disengaging automation and flying manually to maintain basic skills (item 45), while a minority (about 20%) did not. Similar results were found in the present study – 80% endorse disengaging automation to keep their manual skills sharp, and 10% of pilots do *not* endorse this behavior. A substantial minority (13%) in the present study also felt that avoidance of disengaging automation helped to maintain safety (item 42). Additionally, Sherman, Helmreich, and Merritt reported that most (79%) U.S. pilots ensured that the other crew member acknowledged their FMS inputs (item 49), while a small number did not endorse this safety-enhancing practice. The present study essentially replicated this finding – 83% report ensuring that the other pilot acknowledges their FMS inputs, and 8% report not doing this.

In general, examination of the categorical responses in the present study supports the view that there are potential threats to safety in U.S. pilots' attitudes toward automation management. The next analyses are devoted to exploring the latent structure of automation management attitudes, and investigating whether there are differences in automation attitudes as a function of organization, fleet, and respondent demographics.

**Results: Automation item scale development**

The correlation matrix of 20 automation management items was subjected to unconstrained PAF, with oblique rotation of the resultant factor solution. After rotation, the solution yielded three significant (eigenvalues > 1.0) factors, accounting for 30% of the variance. The eigenvalue of the fourth factor was 0.52. Consequent analysis using orthogonal rotation indicated that this factor structure was stable. The obtained factors were identical across rotation methods, although the loading values differed slightly. Because the fourth factor did not possess a sufficiently large eigenvalue, it was not considered further in the factor analyses. The items associated with this factor (as well as other items not loading onto the three retained factors) are addressed in greater detail below.

Examination of the loadings obtained from oblique rotation of the solution showed that nine items loaded significantly (i.e., > .45) onto the three retained factors. Three items loaded onto the first factor, four onto the second factor, and two onto the third factor. Reliabilities for all three factors were good for scales of short length. Cronbach alpha for the first factor was .69; for the second factor, .63; for the third factor, .65. Table 17 contains the factor loadings for the oblique solution.

Table 17. Results of Factor Analyses, Automation Management Items.

| Item | *F* 1 | *F* 2 | *F* 3 | *F* 4 | Λ | Θ |
|---|---|---|---|---|---|---|
| 35. I prefer flying automated aircraft | **.68** | | | | **.71** | **.50** |
| 41. I look forward to more automation | **.67** | .10 | | | **.69** | **.53** |
| 46. Automation leads to safer operations | **.59** | | | | **.66** | **.57** |
| 39. Concerned that automation will cause loss of flying skills | -.35 | | .10 | .23 | | |
| 51. Automation should be used at crew’s discretion | | **-.53** | | | **.42** | **.83** |
| 45. I maintain flying proficiency by disengaging automation | | **-.50** | .18 | -.12 | **.47** | **.78** |
| 42. Pilots should avoid disengaging automation | .19 | **.47** | | | **-.52** | **.73** |
| 53. I try to use automation as much as possible | .42 | **.46** | .12 | | **-.65** | **.57** |
| 48. My company expects me to always use automation | -.14 | .42 | .21 | | | |
| 50. Feel free to select level of automation at any given time | .29 | -.37 | .14 | -.11 | | |
| 37. Effective crewmember always uses the automation | .31 | .36 | .13 | -.21 | | |
| 44. Automated cockpits require more verbal communication | | | **.65** | .11 | **.95** | **.09** |
| 47. Automated cockpits require more cross-checking | | | **.64** | | **.50** | **.75** |
| 52. Flying automated aircraft alters transfer of information | | | .42 | .24 | | |
| 49. I make sure other pilot acknowledges my FMC changes | | | .30 | -.20 | | |
| 43. There are FMC modes and features I don’t understand | | | | .61 | | |
| 40. Easy to forget how to do FMC ops that are not done often | | | | .55 | | |
| 36. I can rapidly access FMC info in abnormal conditions | .25 | | | -.49 | | |
| 54. Difficult to know what FMC ops others are performing | -.11 | | .11 | .44 | | |
| 38. Better to avoid FMC reprogramming in high workload | -.12 | -.17 | .16 | .18 | | |

*Note.* *F* = factor loading for PAF with oblique factor rotation; Λ = CFA lambda; Θ = CFA theta. Items in CFA are in boldface. For PAF, items with loadings < |.10| are not shown.
*Note.* CFA for two target factors yielded a GFI = .96; $\chi^2$ = 135.83 (24 *df*), $p < .001$; $\chi^2/df$ = 5.66.

Inspection of the items comprising the first factor revealed that all three items – “I prefer flying automated aircraft”, “I look forward to more automation”, and “Automation leads to safer operations” – measured pilots’ general preference for automated aircraft. The factor was named “Automation Preference”; higher scores indicate a greater preference for automation on the flight deck. This scale and all other scales are unit-weighted; scores on this scale can range from 3 to 15.

The second factor consisted of items reflecting pilots’ perceived freedom to determine when and how automated systems should be utilized. The items are “Automated systems should be used at the crews’ discretion”, “I regularly maintain flying proficiency by disengaging automation”, “In order to maintain safety, pilots should avoid disengaging automated systems", and "I try to use automation as much as possible during flight operations". The scale was named "Automation Discretion". Higher scale scores indicate increased endorsement of an autonomous, discretionary approach to automation use (the third and fourth items are reverse-scored). Scale scores can range from 4 to 20.

The third factor consisted of items measuring the degree to which pilots recognize the need for increased intra-crew communication and coordination regarding activities on the automated flight deck. The items loading onto this scale were "Automated cockpits require more verbal communication between crewmembers" and "Automated cockpits require more cross-checking of crewmember actions". The factor was named "Recognition of Communication Effects". Higher scale scores indicate greater awareness of the need for increased communication and coordination when using automation. Scores can range from 2 to 10.

Cross-validation of the derived factors using the remaining half of the sample indicated that the factor structure was a good fit to the data. CFA with freely estimated factor correlation specifying the three target factors yielded a $\chi^2/df$ of 5.66 ($\chi^2$ = 135.83 [24 *df*], $p < .001$), and a GFI of .96. Lambda and theta values are also in Table 7.2 above. The factors were retained and utilized in subsequent analyses. Table 18 contains the factor intercorrelations.

Table 18. Pearson Product-Moment Intercorrelation Matrix for Scales and Selected Individual Items.

| | APref | ADiscret | RCE | Item 37 | Item 38 | Item 39 | Item 49 | Item 50 |
|---|---|---|---|---|---|---|---|---|
| APref | - - | -.33** | -.09** | .39** | -.21** | -.38** | .07* | .19** |
| ADiscret | | - - | .08* | -.40** | .21** | .15** | -.06* | .18** |
| RCE | | | -- | -.03 | .18** | .12** | .17** | .06* |
| Item 37 | | | | - - | -.17** | -.21** | .13* | .06* |
| Item 38 | | | | | - - | .16** | -.02 | -.02 |
| Item 39 | | | | | | - - | -.08* | -.15** |
| Item 49 | | | | | | | - - | .13** |
| Item 50 | | | | | | | | - - |

*Note.* APref = Automation Preference scale. ADiscret = Automation Discretion scale. RCE = Recognition of Communication Effects scale. Item 37 = "The effective crew member always uses the automation tools provided". Item 38 = "When workload increases, it is better to avoid reprogramming the FMC". Item 39 = "I am concerned that the use of automation will cause me to lose flying skills". Item 49 = "I make sure the other pilot acknowledges programming changes I make in the FMC". Item 50 = "I feel free to select the level of automation at any given time".
*Note.* *$p < .05$, **$p < .001$.

Inspection of the factor intercorrelations reveals an interesting relationship: Automation Preference is negatively correlated with Automation Discretion ($r = -.33$ (1524 *df*), $p < .001$). This implies that pilots who express a high preference for automation also show lower endorsement of discretionary automation use, as well as lower endorsement of disengaging automation to maintain manual flying skills. This replicates Sherman, Helmreich and Merritt's (in press) similar finding across 12 different nations.

**Items not included in scaling**. Eleven items did not make the 'empirical cut'. Three of these items were intended to tap respondents' level of concern with potentially deleterious aspects of FMS use, such as forgetting little-used operations, losing manual flying skills, and trying to carry out FMS operations during abnormal conditions. They are item 36, "Under abnormal conditions, I can rapidly access the information I need in the FMC"; item 39, "I am concerned that the use of automation will cause me to lose flying skills"; item 40, "It's easy to forget how to do FMC operations that are not performed often"; and item 43, "There are modes and features of the FMC that I do not fully understand").

Three items were intended to indicate respondents' recognition of increased communication and coordination needs on the automated flight deck (item 49, "I make sure the other pilot acknowledges programming changes I make in the FMC"; item 52, "Flying highly automated aircraft alters the way crew members transfer information"; and item 54, "It's difficult to know what FMC operations the other crewmember is performing").

Also, three items had been written to indicate respondents' level of discretion in automation use (item 37, "The effective crew member always uses the automation tools provided"; item 38 "When workload increases, it is better to avoid reprogramming the FMC"; and item 50, "I feel free to select the level of automation at any given time"). Finally, one item (item 48, "My company expects me to always use automation") was intended as an indicator of respondents' perceptions regarding company policy; as such, it was not expected to scale. It was included in exploratory factor analyses because of the chance that it might have significantly covaried with responses to automation discretion items.

While it cannot be precisely determined why items did not behave as intended, some educated guesses can be made about several items. Upon further inspection, it became clear that three of the 'concern' items (items 36, 40, and 43), while originally intended to measure respondents' awareness that automation can be an opaque and sometimes clumsy means of accomplishing tasks, could also be answered from a knowledge- or skill-based standpoint. In other words, agreement with "there are modes and features of the FMC that I do not fully understand" can be viewed as a healthy appreciation of the opacity of the FMS; however, agreement could also imply a lack of skill with the FMS. Additionally, disagreement with the statement could indicate that the respondent did not have a healthy appreciation of FMS opacity, *or* that the respondent was in fact quite skilled with the FMS. Put simply, these three items confound concern with automation use and knowledge of systems. The remaining concern item "I am concerned that the use of automation will cause me to lose flying skills" does not appear to suffer from this confound, and will be described as a stand-alone indicator of automation concern.

Of the three unscaled items written to measure respondents' awareness of increased communication and coordination requirements when using automation, one item (item 54, "It is difficult to know what FMC operations the other crewmember is performing") suffers from the same confound described above. It is impossible to ascertain whether agreement with this item denotes an appreciation of the increased communication demands on the automated flight deck, or a deficit in communication skill on the part of the respondent or the other crewmember. The two other items in this grouping seem to be relatively 'pure' indicators of the intended construct. Of these two, item 49 ("I make sure the other pilot acknowledges programming changes I make in the FMC") references a specific behavior; it was chosen for individual analysis.

Finally, the three items reflecting respondents' discretion in automation use seem to be fairly clear indicators of the intended construct; they are also be utilized in item-level analyses. In sum, five individual, unscaled items were identified as suitable for inclusion in further analyses. Three of these items are conceptually related to the Automation Discretion scale, one is conceptually related to the Recognition of Communication Effects scale, and one is a stand-alone indicator of respondents' concern with the potential for automation use to cause loss of manual flying skills. The inter-item and item-scale correlations are also in Table 17 above.

**Results: Correlations of automation management scales with experience, indicators of organizational policy, and Uncertainty Avoidance Values**

The three automation management scales (along with their component items), four single item indicators of scale constructs, and one single item indicating concern with loss of manual flying skills were correlated with flying experience, two items measuring pilot perceptions of company policy toward automation use, and scores on the Uncertainty Avoidance Values scale (a measure of individuals' need to avoid uncertain, ambiguous situations with increased proceduralization and formalization of work tasks). Overall, there were relatively few high correlations. Table 19 contains the correlations between these sets of variables.

**Demographics.** There were no meaningfully significant correlations between flight hours in pilots' current aircraft and the automation scales or single item indicators; i.e., although some correlations reached statistical significance, none were greater than |.10|. Total flight hours were slightly negatively correlated with Automation Preference scores ($r$ = -.10 [1446 *df*], $p$ < . 001), suggesting that older pilots prefer automation slightly less than do younger pilots. Recalling the finding in Chapter Five showing that older pilots evaluated their transition training more negatively, it is not surprising that these same pilots would also show a lower preference for automation.

Total flight hours were slightly positively correlated with Recognition of Communication Effects scores, $r$ = .10 (1592 *df*), $p < .001$, suggesting that older, more experienced pilots show slightly greater recognition of increased communication demands on the automated flight deck. Given older pilots' more negative evaluations of training and wariness toward automation in general, it is also not surprising that a small relationship exists between overall experience and the recognition of the need to increase intra-crew communication and cross-checking of each other's actions. Total flight hours were not correlated with the single item indicators.

Table 19. Correlation of Three Automation Management Scales, their Component Items, and Five Unscaled Automation Management Items with Two Experience Variables, Two Indicators of Perceived Company Policy, and Uncertainty Avoidance Values Scale Scores.

| Scales and items | Tot hrs | Curr hrs | Item 32 | Item 48 | UAV |
|---|---|---|---|---|---|
| Automation Preference | -.10** | .04 | .18** | -.05* | .11** |
| 35. I prefer flying automated aircraft | -.11** | .05 | .12** | -.06* | -.06* |
| 41. I look forward to more automation | -.08* | .04 | .11** | -.01 | .11** |
| 46. Automation leads to safer operations | -.06* | .01 | .18** | -.05* | .07** |
| Automation Discretion | -.07* | .06* | .10** | -.24** | -.15** |
| 42. Pilots should avoid disengaging automation | .07* | -.03 | -.03 | .17** | .10** |
| 45. I maintain flying proficiency by disengaging automation | .01 | .08* | .08* | -.13** | -.08* |
| 51. Automation should be used at crew's discretion | -.04 | .05 | .19** | -.17** | -.07* |
| 53. I try to use automation as much as possible | .08* | -.01 | .04 | .20** | .16** |
| Recognition of Comm. Effects | .10** | -.02 | -.06* | .12** | .04 |
| 44. Automated cockpits require more verbal communication | .05 | -.05 | -.03 | .07* | .03 |
| 47. Automated cockpits require more cross-checking | .13** | .02 | -.08* | .14** | .04 |
| 37. Effective crewmember always uses the automation | .03 | -.01 | -.03 | .18** | .20** |
| 38. Better to avoid FMC reprogramming in high workload | .04 | .02 | .01 | .01 | -.02 |
| 39. Concerned that automation will cause loss of flying skills | .02 | -.03 | -.13** | .12** | .03 |
| 49. I make sure other pilot acknowledges my FMC changes | .06 | -.07* | .05 | .08* | .09** |
| 50. I feel free to select level of automation at any given time | .01 | .06* | .25** | -.15** | .06* |

*Note.* Tot hrs = total flight hours. Curr hrs = flight hours in current aircraft. Item 32 = "During transition training, I was taught that the decision to use or not use automation on this aircraft is up to me". Item 48 = "My company expects me to always use automation". UAV = Uncertainty Avoidance Values scale.
*Note.* *$p < .05$, **$p < .001$.

**Organizational policy.** Responses to the organizational policy perception item "During transition training, I was taught that the decision to use or not use automation on this aircraft is up to me" were positively correlated with Automation Discretion, $r$ = .10 (1669 *df*), $p < .001$. Bolstering this finding was the significant correlation between this measurement of company policy perceptions and the unscaled indicator of automation discretion, "I feel free to select the level of automation at any given time" ($r = .25$ [1672 *df*], $p < .001$). Taken together, these relationships suggest that pilots who perceive that company policy allows them greater flexibility to utilize (or not utilize) automation, also endorse greater discretion in their personal use of automation.

Supporting this assertion is the negative correlation between the item "My company expects me to always use the automation" and scores on the Automation Discretion scale ($r = -.24$ [1670 *df*], $p < .001$), as well as the positive correlation between this company expectation item and responses to the unscaled indicator of automation discretion, "The effective crewmember always uses the automation tools provided" ($r = .18$ [1675 *df*], $p < .01$). These relationships further suggest that perceptions that the company expects maximal automation use are associated with a less autonomous, discretionary use of automation during flight operations.

Both company policy perception items were significantly correlated with the single item indicator of automation concern, "I am concerned that the use of automation will cause me to lose flying skills". Responses to "During transition training, I was taught that the decision to use or not use automation is up to me" were significantly correlated with response to the automation concern item, $r = -.13$ (1674 *df*), $p < .001$, suggesting that pilots who felt that the company supported discretionary automation use were less concerned with losing their manual flying skills (presumably because they felt more comfortable preventing loss of flying skills by disengaging automation and hand-flying). This supposition is supported by the positive correlation between the concern item and responses to "My company expects me to always use the automation" ($r = .12$ [1509 *df*], $p < .001$). Finally, "During transition training, I was taught that the decision to use or not use automation is up to me" was positively correlated with Automation Preference scores, $r = .18$ (1525 *df*), $p < .001$, suggesting that pilots who felt their company encouraged discretionary automation use also show a greater preference for automation.

Overall, these correlations lend support for the systems view of aviation – i.e., they suggest that perceptions of organizational policies do in fact relate to pilots' reported automation management practices. Put more simply, these data suggest that perceptions of company policy may have considerable power to determine how pilots feel they should use (and not use) automation.

**Uncertainty avoidance.** Although Merritt (1996) reports high reliability for the Uncertainty Avoidance Values scale (conceptually based on Hofstede's [1980] Uncertainty Avoidance scale) when examined across nations (Cronbach alpha was .89 for the six-item composite when measured at the national, as opposed to individual, level), relatively little is known about the scale's intra-national properties. To investigate this, Cronbach's alpha was calculated for the six items comprising this scale. The scaled items include item 59, "The organization's rules should not be broken – even when the employee thinks it is in the organization's best interests"; item 73 "Written procedures are necessary for all in-flight situations"; item 79 "[How important would it be to you to] observe strict time limits for work projects?"; item 81 "[How important to...] know everything about the job, to have no surprises"; the reversed item 82, "[How important to...] have a changing work routine within new, unfamiliar tasks?"; and item 83, "[How important to...] find the truth, the correct answer, the one solution?"

Within the U.S., reliability was .49, which was low, but still acceptable. The reliability coefficient most likely was low because the items, representing widely shared intra-cultural values, show some restriction of range, which can negatively affect Cronbach alpha calculations (Merritt & Helmreich, 1997). Consequent recalculations of alpha while eliminating items one-by-one did not appreciably improve reliability, so it was decided to utilize the scale as constructed by Merritt.

Scores on the Uncertainty Avoidance Values scale were slightly positively correlated with Automation Preference scores ($r = .11$ [1505 *df*], $p < .001$), and slightly negatively correlated with Automation Discretion scores ($r = -.15$ [1506 *df*], $p < .001$). Additionally, Uncertainty Avoidance Values was positively correlated with the unscaled indicator of automation discretion, "The effective crewmember always uses the automation tools provided" ($r = .20$ [1509 *df*], $p < .01$). These results suggest that the theoretical expectation described earlier may be tenable – pilots with a high need to avoid uncertain, ambiguous situations support a less discretionary approach toward use of automation. That is, they feel that it should be used more, and show less endorsement for disengaging it. Also in line with the theoretical expectation, higher uncertainty avoidance is associated with greater preference for automation.

**Results: Scale score comparisons using ANOVA techniques**

Responses to the Automation Preference scale were studied in a four-factor (Airline X Fleet X Previous Automation Experience X Position) ANCOVA, with total flying hours utilized as a covariate (recall that total flying hours was significantly negatively correlated with Automation Preference scores). Responses to the Automation Discretion and Recognition of Communication Effects scales were analyzed in two four-factor ANOVAs, with no covariates. Crew position was utilized as a factor in these analyses because pilots' attitudes toward management of flight tasks can sometimes differ depending upon the team role they take on (i.e., leader or subordinate) in the cockpit (Gregorich, Helmreich, & Wilhelm, 1990).

There were no significant main effects or interactions for the Automation Preference ANCOVA, and the covariate was not significant. There were no significant main effects or interactions for the Automation Discretion ANOVA. There was only one significant main effect for the Recognition of Communication Effects ANOVA; pilots with previous automation experience were slightly more aware of the need to increase communication and cross-checking when using automation, $F = 4.03$ (1,1573 *df*), $p < .05$. The mean score for those with previous automation experience was 8.09 (SD = 1.78); for those without previous experience, the mean was 7.96 (SD = 1.73). It should be noted, however, that power was < .80 for all effects and interactions, for all three analyses. This calls into question the likelihood of finding a significant effect (or effects) if in fact they existed. Generally, however, these null findings complement the findings of Sherman, Helmreich, and Merritt (in press) showing only relatively small cross-organization differences (that is, within a given nation) on automation attitudes.

## Summary

Inspection of the item-level categorical responses indicated that overall, a substantial minority of pilots do not endorse automation management attitudes and practices that may lead to increased safety margins. These findings replicate earlier investigations using both U.S. pilots and pilots from 11 other nations.

Correlational analyses indicated that generally, older pilots preferred automation slightly less than did younger pilots; older pilots also showed slightly more recognition of the increased need for communication on the automated flight deck. Perceptions of company automation use policy were substantially correlated with a number of automation management measures (both scales and individual items). In general, respondents who perceived that their company expected automation use and did not teach a discretionary approach to automation management, showed less endorsement of discretionary automation use, more concern with the potential of loss of manual flying skills associated with automation use, and lower preference for automation. This offers considerable support for the explanatory power of the systems view of aviation.

Finally, Uncertainty Avoidance Values scores were slightly negatively correlated with measures of discretionary automation use, and positively correlated with Automation Preference. This indicates that individuals with a high need to avoid ambiguous, uncertain situations may also endorse increase use of automation as a means of reducing perceived uncertainty.

## DISCUSSION

This section discusses implications of the potential safety threats identified in this study, and proposes some ways of ameliorating these threats.

### Implications and Recommendations Regarding Equipment Evaluations

Examination of automated systems evaluations yielded several interesting findings. A number hold the potential to lower safety margins. For example, the ease with which crews gather information from the FMS can certainly affect the safety of flight. Recall for the moment that flight plans are frequently changed by air traffic controllers *during flight*, particularly in the busy areas around airports when many aircraft are climbing and descending in close proximity. Entering or extracting information from a balky or difficult to use FMS while close to other aircraft (and the ground!) can reduce the time and attention crews devote to other tasks, such as scanning for other aircraft and configuring the aircraft for approach and landing.

In a similar vein, flight deck automation that provides insufficient feedback regarding mode changes, system states (in both normal and abnormal conditions), and thrust levels, essentially forces crews to attend to the automation instead of to the management of the overall flight. This is a clear example of a safety threat. To illustrate this, an analogy based on the view of automation-as-crewmember can be drawn. It has been shown in studies of three-person air crews that when one crew member performs team functions (but not necessarily technical functions) inadequately, the remaining crew members increase team performance levels in an attempt to compensate (Chidester & Foushee, 1989). In other words, in the presence of a crew member who is not adequately communicating intentions and plans, the others increase their level of teamwork and take additional care to elicit the needed information from the inadequately-performing crewmember. This has the effect of increasing the workload for the adequately performing individuals, because two are doing the (team)work of three.

If a particular automated system, whether charged with annunciation of thrust levels or lateral and vertical navigation modes, doesn't adequately communicate *its* intentions and plans, a similar pattern of team performance can occur. To meet the challenge posed by this situation, air crews will de facto increase their performance (and often their workload) to compensate for the inadequately-performing crewmember (i.e., the automation) by taking extra care to elicit needed information from the automation and from each other. However, safety margins will be reduced, because, as in the all-human flight deck, two will be doing the teamwork of three[17].

Because all human endeavors are prone to error, and error (all things being equal) is more likely to occur when performance demands are higher, it becomes clear that when automation is not a 'team player', the likelihood of error commission (and therefore, accidents) increases. Incidents and accidents can result from a number of specific automation vulnerabilities. For example, an FMS interface where navigation information is presented to air crews according to poorly-designed decision rules can lead to accidents similar to the earlier-described B757 crash near Cali, Columbia. A thrust management system with highly autonomous and opaque mode changes can lead to accidents like the 1988 crash of an A320 at Habsheim, France. In this accident, the pilots did not notice that the aircraft's thrust management system had undergone a mode change, and neglected to manually increase thrust during a low, slow pass near an air show bleacher (Dornheim, 1995). Finally, poorly designed methods of manual override can leave air crews unsure whether the pilots or the FMS is controlling the aircraft; this has been a factor in a number of incidents and accidents (Funk, Lyall, & Riley, 1995).

It should be recalled that the equipment evaluation differences described in the present study are relative, not absolute. It is difficult to ascertain the degree to which lessened thrust awareness or lower ratings of crew alerting and advising systems negatively affect performance on each specific aircraft type. Therefore, although McDonnell-

[17] This is also a subtly different take on Parasuraman and Riley's view of the automation as a proxy for the designer. It suggests that designer's *assumptions about team processes*, as well as their gross errors, are represented on the flight deck (or in the control tower, train cab, bridge, or operations room).

Douglas and Airbus aircraft receive lower ratings on several dimensions, aircraft from these manufacturers cannot be said to be to be less safe in an absolute sense.

Although the findings detailed here are relative, the present study does reinforce the overall conclusion that many issues in the crew-automation interface require additional study. If these potential safety threats are to be addressed, it is vital that the work not stop with the present study. The fact that the opinions of nearly 2,000 U.S. commercial pilots are represented here does not absolve researchers, manufacturers, and regulatory bodies from pursuing further lines of inquiry. It is not sufficient to know that U.S. pilots feel thrust awareness is slightly lower in certain aircraft; further work must be done to ascertain *why* this is so, and *what* can be done to ameliorate lowered awareness of thrust levels in specific aircraft.

While further work in this area is necessary, the results of the present study do offer support for some general guidelines. The present study strongly suggests that systems designers, airframe manufacturers (indeed, any purveyors of automated solutions) should follow a general set of recommendations before automating a process. These recommendations can be summarized as follows:

1. the human operators should retain sufficient command of the operation,
2. the operators should remain sufficiently involved (i.e., play an active role) such that they can recognize and ameliorate deviations from intended conduct of an operation,
3. operators should be kept appropriately informed regarding automated systems behavior,
4. the systems' behavior should be sufficiently predictable,
5. the systems should help prevent error by monitoring the operators, and
6. the systems should help guard against error by enhancing crew cross-monitoring tasks to the greatest extent possible (Sherman, Helmreich, & Hines, 1997, adapted from Billings, 1997).

This last requirement makes explicit the need for automated systems to contribute positively to the maintenance of team performance.

Similar to the above recommendations, ALPA (1996), Billings (1997), and the NAS (in press) all recommend asking these types of questions before an automated solution is applied to a task: Why is this function being automated? Will automating the new function preserve operator involvement, and enhance operator teamwork and situation awareness? Would *not* doing so improve the operators' involvement, teamwork, and situation awareness?

Using the results of the present study to guide further inquiry, and following the guidelines presented above, may aid both manufacturers and end-users in the quest to address the automation-crew vulnerabilities identified by ALPA (1996), the FAA (1996), and others. It should be recalled, however, that although the differences in evaluations across aircraft types and airframe manufacturers are interesting in and of themselves, pilots' evaluations of the automated equipment cannot be considered in isolation. They must be considered in conjunction with their training experiences.

## Recommendations Regarding Training Evaluations and Automation Management Attitudes

Even the most well designed automated system is vulnerable to misuse, disuse, and abuse (Parasuraman & Riley, in press). Though some of these non-optimal patterns of use are undoubtedly due to deficiencies in systems design, some may be due to deficiencies in training and organizational guidelines for use of systems.

If operators of automation are not given appropriate training in the use of automation, or the organization does not support the lessons imparted in appropriate training with explicit, empirically informed guidelines for use, safety margins can be reduced. This does not just apply to the aviation industry. The NTSB (1997), in a report on a recent maritime accident, cited the crew's overreliance on an automated navigation system and subsequent failure to monitor the system as causing a 17-mile course deviation, resulting in the grounding of a cruise ship near Nantucket, Massachusetts. One of the primary antecedents of the accident, according to the NTSB, was inadequate training of the ship's crew in the use of the automation. The Safety Board also reported that the *sole* (not just primary) means of gaining proficiency in use of automated bridge equipment for this company's fleet had been on-the-job training[18]. No formal instruction had been given to the crew regarding how to operate the automation. Additionally, no formal guidelines regarding when to use automation, and how to monitor it, had been promulgated.

This is an extreme example. Generally speaking, training for operation of aircraft automation is much more extensive than this. However, economic necessities almost always dictate that pilots transitioning to an automated aircraft be returned to the line as soon as possible; thus there is considerable pressure to fit as much course material as possible into the limited training period. In general, most aircraft training programs walk a fine line between imparting a level of skill and knowledge sufficient to operate the aircraft, and returning the pilot to the line quickly (Sherman & Helmreich, in press). When the high cost of dealing with training failures is also considered, it becomes apparent that there is the potential for training to only marginally prepare some pilots for the line.

The findings that overall, roughly 25% of crews did not gain an adequate understanding of the FMS during training, and felt unprepared for line flying at the end of training, bear this out. These findings are also confirmations of the concerns voiced by ALPA (1996), the ATA (1996), and the FAA (1996), and represent potential threats to safety. The considerable variability in reported training efficacy across airlines, across airlines within single aircraft types, and within airlines across aircraft types is simultaneously a source of concern and a source of optimism.

The variability across these factors raises concerns because it implies that there are widely differing standards, expectations, and outcomes across fleets and airlines. In an industry historically committed to standardization, these findings may be viewed with surprise and dismay. However, the findings should also generate considerable optimism – the fact that some fleets and airlines earn better training efficacy ratings than do others strongly implies that the training vulnerabilities described here are *tractable*.

Although confidentiality agreements preclude revealing more specific information regarding the fleets and airlines represented in the present study, each airline represented in the study will be afforded a much more comprehensive look at their own pilots' responses (at the fleet level). Careful examination (and adaptation) of training practices and procedures carried out in higher-rated fleets and airlines should help to improve training for lower-rated fleets and airlines. In this way, the data that were not reported here may also aid in the effort to identify and ameliorate safety threats.

The results regarding free-play efficacy also point to one potential point of leverage in the efforts to bolster air crew's knowledge, proficiency and comfort with automated systems. As part of investigation for a separate study concerned with free-play, it was found that there is considerable demand by pilots for more and better free-play devices (Sherman & Helmreich, in press). The training departments of several airlines indicated that, in response to pilot demand, they are presently deploying PC-based part-task CDU/FMS free-play devices in some of their fleets. Another airline recently accomplished this PC-based device deployment, and has also made available for extra-curricular use cockpit procedures trainers and flight management systems trainers. In general, there is high demand for access to these devices (one training department manager reports that pilots "practically kill for" access to these devices [Ratwatte, 1997]).

Implementing increased free-play opportunities in an effort to bolster FMS knowledge, proficiency, and comfort may be particularly effective because it leverages individuals' seemingly boundless sense of curiosity, as well as the deep sense of professionalism inherent to the pilot ranks. This is borne out by a recent article in *Aviation Week and Space Technology* indicating the considerable amount of extra-curricular FMS training and discussion occurring within the aviation industry, citing as evidence the popularity of third-party FMS manuals, and Internet-based discussion groups devoted to flight deck automation (Dornheim, 1996b).

As a general rule, providing increased access to free-play so pilots can bolster and exercise their automation-related knowledge and skills may prove to be a useful means of combating inadequate and inaccurate

[18] Although the NTSB and others see on-the-job training as a valuable means of gaining proficiency, it should probably not be used as the primary means of training for use of automation.

conceptions of flight management systems behavior. Increased access to free-play may also prove relatively inexpensive to implement, especially in contrast to the cost of high-fidelity simulators.  Finally, as airlines move their fleets to Advanced Qualification Program-based standards, the proficiency-based nature of free-play learning will probably make it even more attractive to training departments.

At this point several caveats must be offered. This design of this study did not allow ascertainment of *how often* pilots were able to practice, or *what type* of device they utilized. Many different devices can be used for free-play; these devices can differ in the flexibility and operational fidelity they offer. Also, the present study could only address the relationship between free-play ratings and training ratings; the question remains as to whether free-play actually leads to improved training and line performance outcomes[19].

Finally, it should be mentioned that training departments must take care to provide high-quality instruction in FMS logic and operation, concomitant with these extra-curricular learning sessions. As Woods (1997) points out, peer learning can be a powerful tool, but peer consensus cannot substitute for expert knowledge and skilled pedagogy. Despite these shortcomings and caveats, though, the results still suggest that free-play may be an especially useful training tool for automated fleets.

However, extra-curricular methods of bolstering FMS knowledge, comfort and proficiency can move the industry only so far. As the FAA (1996) states, it is incumbent upon training departments to improve flight training for automated aircraft on several fronts. In summary, training departments must rise to the challenge of providing pilots with 1) the technical knowledge to understand the automation; 2) the wisdom to develop adequate judgment regarding automation use; and 3) training experiences relevant to the operating environment.

The concept of “training to the threat” (i.e., providing training that addresses problems found in actual operations) deserves particular attention here. The ATA (in press) points out that training for automated aircraft (in fact, for most commercial aircraft) too often takes the form of a ‘fill-in-the-blocks’ exercise. In these types of training programs, pilots are instructed in a set of FAA-mandated procedures and topics, with little attention to the particular problems and challenges faced by the individual fleet, and the airline as a whole. This approach does not address specific threats that an airline or fleets within an airline typically encounter in their particular operating environment.

One major U.S. airline with substantial operations in Latin and South America has recently implemented a training program designed to teach pilots how to react to uncertain situations when flying into mountainous areas with limited air traffic control and airport services. Portions of this training also stress the importance of not becoming overly reliant upon the navigation database found in the flight management computers of certain aircraft (American Airlines, 1996), as research suggests that aircrews have a tendency to over-rely on automation to “perform tasks and make decisions for them, rather than using the aids as one component of thorough monitoring and decision-making processes” (Mosier, Skitka, & Heers, 1995, p. 222). This can be especially dangerous when flying into and out of mountainous or hazardous areas. The program described above is a good example of providing training commensurate to the threats posed by the interaction of automation and environmental factors.

Environmental conditions, however, are not the only variables to consider when attempting to identify specific threats to safety. As the present study shows, threats related to automation use can come from elsewhere within the systems perspective, such as individuals’ values and attitudes, and airlines’ organizational cultures.

Examination of categorical responses to individual automation management items in this study provides useful but disturbing information. The finding that a considerable number of pilots did not see the value of avoiding reprogramming the FMC when workload increases, did not endorse disengaging automation to keep their manual skills sharp, and did not report ensuring that the other pilot acknowledges their FMS inputs, are unsettling findings. In an absolute sense, it suggests that some pilots’ automation use judgement may require remediation. Also, the fact that there is not greater consensus among pilots regarding these fundamental, safety-enhancing behaviors implies that, for any given flight, individuals comprising the air crew may hold drastically different conceptions of how the automation should be used.

---

[19] Happily, this is an empirical question that can and should be addressed in future research.

This considerable divergence of opinion toward aspects of automation management is an indication that crews would be well advised to make explicit their predilections and preferences regarding management of the FMS. Ideally, sharing this type of information would first occur in the initial crew briefing at the beginning of a trip, where crew members first meet, and usually articulate their expectations for performance of flight duties. Unfortunately, previous research (e.g., Helmreich, Hines, & Wilhelm, 1996) has shown that a fairly large minority of crews conduct below-standard briefings. Thus, for crews of automated aircraft, the challenge to conduct adequate briefings is compounded by the need to adequately address the issues surrounding management of the automation.

These observations also present a challenge to the organizations themselves. Organizational culture's effects on individual and team behavior can be profound, although the data presented here indicate relatively small relationships between perceptions of organizational values regarding automation use and attitudes toward aspects of automation management. Still, it is fair to say that expectations and norms for flight crew performance are to a large extent inculcated by the organization. It falls to the organization to promote the expectation that automation management requires greater attention in briefings and during flight.

These assertions lead to the conclusion that it is not enough for a training department to teach automation knowledge, management and judgment. In order to bring about optimal automation use, the organization as a whole must actively reinforce the lessons imparted in training, by creating *and* effectively disseminating guidelines for automation use that complement the lessons of well-designed transition training (FAA, 1996). As Degani and Wiener (1994) observe, this requires that an organization strive for consistency between operating philosophies, policies, and procedures, as a means of bringing about intended flight crew practices. Lamentably, this is not the case in many organizations today (Helmreich, Hines, & Wilhelm, 1996). Even among organizations that *have* striven to articulate sensible guidelines for automation use, a considerable proportion of pilots perceive that maximal automation use is incumbent upon them (ATA, in press).

The importance of the organization in the effort to promote safe automation use strategies (and the potential of the organization to undermine safe automation practices) probably cannot be overstated. An excerpt from a letter included with a survey returned by one pilot with 20 years of airline experience illustrates this point:

> I personally have always clicked off the automation in order to keep my personal flying skills sharp. This practice, until recently, has been in violation of the company written procedures and training philosophy…as a result, [among other pilots] I have observed lowered basic skills, including roughness on the controls, lowered situational awareness, and falling behind the airplane when hand flying.

This comment is anecdotal, yet it speaks volumes about the importance of achieving congruence between pilots' (indeed, any automation operators') needs and predilections, and the organization's philosophies and guidelines. The results of the present study may aid in the effort to achieve this congruence by serving as springboards from which to launch further inquiry.

## APPENDIX A: THE UNIVERSITY OF TEXAS AVIATION AUTOMATION SURVEY

As part of NASA and FAA sponsored research, we are collecting flight operations data from pilots and other flight personnel in a variety of cultural settings around the world. This questionnaire is part of a world-wide study aimed at understanding attitudes toward automation use and flight management attitudes in different countries. All data are strictly confidential and results will be presented only at the group level. No individual feedback will be given to management, so feel free to express your opinions. Your participation in the study is valued and appreciated.

### I. Background Information

Airline ________________

How many years have you been with this organization as a pilot? _________

How many years employed within the commercial aviation industry overall? __________

Initial flight training background (check one) Military ____________ Civilian _____________

What is your nationality? _________________ Nationality at birth (if different than your present nationality) ______________

Gender (M or F) _______

Total flying hours (including time as second officer, if any) __________

Airplane you are currently flying (type and series): __________

Hours in this type: __________

*T*ype-rated in this a/c? (Y or N) __________

| Crew Position: | Status: |
|---|---|
| _______ Captain | _______ Line Pilot |
| _______ First Officer | _______ Instructor/Training Pilot/Supervisor |
| _______ Second or Relief Officer (long haul) | _______ Check Airman/Licensing Examiner/Designator |
| | _______ Management |
| | _______ Other _________________ |

| Other automated aircraft you are current in: | Hours in this type: | Highest position attained: |
|---|---|---|
| _________________________ | ____________ | _____CA _____FO _____SO/Relief |
| _________________________ | ____________ | _____CA _____FO _____SO/Relief |
| _________________________ | ____________ | _____CA _____FO _____SO/Relief |

| Other commercial aircraft types you have flown: | Hours in this type: | Highest crew position attained: |
|---|---|---|
| _________________________ | ____________ | _____CA _____FO _____SO/Relief |
| _________________________ | ____________ | _____CA _____FO _____SO/Relief |
| _________________________ | ____________ | _____CA _____FO _____SO/Relief |

### II. Training Background

Please answer these next items with regard to your current aircraft type. If you currently fly more than one type of automated aircraft, answer with regard to the aircraft type you fly most often.

**Transition Training (for aircraft you are currently flying)**

Please indicate approximately how many hours were devoted to each of the following methods of training.

I think there should have been….. (circle a number, or n/a if not applicable)

Classroom _____hours *A lot less* 1 2 3 4 5 *A lot more* n/a

Computer-based _____hours *A lot less* 1 2 3 4 5 *A lot more* n/a

Simulator_____hours *A lot less* 1 2 3 4 5 *A lot more* n/a

Aircraft _____hours *A lot less* 1 2 3 4 5 A *lot more* n/a

Other (specify) _____hours *A lot less* 1 2 3 4 5 *A lot more* n/a

_______________________________________

I was trained by (check one): my airline _____ the aircraft manufacturer _____ other _____

During training, were you able to practice on a CDU/FMS part-task simulator? (yes or no) _____

If yes, how useful did you find this type of training? (choose from the scale below) _____

| A | B | C | D |
|---|---|---|---|
| Not at all useful | Slightly useful | Somewhat useful | Very useful |

**Recurrent Training**

Please indicate how often you receive the types of recurrent training listed below:

| | Every 3 months | Every 6 months | Yearly | Less than yearly | Never or n/a |
|---|---|---|---|---|---|
| Computer-based trng | | | | | |
| LOFT | | | | | |
| Proficiency training | | | | | |
| Systems review | | | | | |

## III. Attitudes Toward Equipment and Training

Please answer the following questions in regard to the type of automated aircraft *you are currently flying*. If you currently fly more than one type of automated aircraft, answer the questions in regard to the aircraft type you fly most often. Please answer by writing beside each item a letter from the scale below.

| A | B | C | D | E |
|---|---|---|---|---|
| Disagree Strongly | Disagree Slightly | Neutral | Agree Slightly | Agree Strongly |

**Equipment**

___ 1. The flight management system (FMS) on this aircraft is easy to use.

___ 2. I am satisfied with the format of the displays on this aircraft.

___ 3. The autothrottle system/computer executes thrust level/airspeed changes smoothly.

___ 4. It is always apparent when an automated system fails.

___ 5. The error messages I receive from the FMS are easy to understand.

___ 6. The FMS in this aircraft provides me with just the right amount of information.

___ 7. The programming procedures of the FMS are well designed.

___ 8. The autothrottle system/computer executes thrust level/airspeed changes accurately.

___ 9. It is very difficult to get the information I need from the FMS in this aircraft.

___ 10. When my inputs are not accepted by the FMS, it is difficult to determine why.

___ 11. I am always aware of changes to thrust level that are commanded by the autothrottle system.

___ 12. If an automated system fails, I understand the nature of the failure quickly.

___ 13. This aircraft was designed to keep the pilot "in the loop" as much as possible.

___ 14. I get insufficient feedback regarding mode changes from the automated systems.

___ 15. It is always apparent when an automated system is doing something other than I expect it to.

___ 16. The error messages I receive from the FMS are helpful.

___ 17. It is difficult to transition from automated to manual flight in this aircraft.

___ 18. When I move the throttles, I expect immediate thrust change to be commanded.

___ 19. I find all modes of the FMS useful.

___ 20. The error messages I receive from the FMS do not help me to solve a problem.

___ 21. I try to use all the modes and features of the FMS.

___ 22. Information I may need from the FMS is easily accessible.

___ 23. I receive useful information about engine thrust levels from the position of the throttles.

___ 24. Information I may need from the FMS is logically presented.

___ 25. The error messages I receive from the FMS are useful in correcting a problem.

**Training**

___ 26. My transition training adequately prepared me for flying this aircraft on the line.

___ 27. This aircraft does not make efficient use of a pilot's basic aviation skills.

___ 28. Recurrent training for my aircraft covers important abnormal procedures that are not encountered very often.

___ 29. I gained an adequate understanding of the FMS during transition training.

___ 30. I learned most of what I need to know about this aircraft when I started flying it on the line.

___ 31. My past aviation experience prepared me for operating this aircraft.

___ 32. During transition training, I was taught that the decision to use or not use automation on this aircraft is up to me.

___ 33. My company's computer-based training for this aircraft is effective.

___ 34. I rely on my basic aviation skills when flying this aircraft.

## IV. Automation Management Attitudes

The following items deal with attitudes regarding flightdeck automation management. Please answer by writing beside each item a letter from the scale below.

| A | B | C | D | E |
|---|---|---|---|---|
| Disagree Strongly | Disagree Slightly | Neutral | Agree Slightly | Agree Strongly |

___ 35. I prefer flying automated aircraft.

___ 36. Under abnormal conditions, I can rapidly access the information I need in the FMC.

___ 37. The effective crewmember always uses the automation tools provided.

___ 38. When workload increases, it is better to avoid reprogramming the FMC.

___ 39. I am concerned that the use of automation will cause me to lose flying skills.

___ 40. It's easy to forget how to do FMC operations that are not performed often.

___ 41. I look forward to more automation.

___ 42. In order to maintain safety, pilots should avoid disengaging automated systems.

___ 43. There are modes and features of the FMC that I do not fully understand.

___ 44. Automated cockpits require more verbal communication between crewmembers.

___ 45. I regularly maintain flying proficiency by disengaging automation.

___ 46. Automation leads to safer operations.

___ 47. Automated cockpits require more cross-checking of crewmember actions.

___ 48. My company expects me to always use automation.

___ 49. I make sure the other pilot acknowledges programming changes I make in the FMC.

___ 50. I feel free to select the level of automation at any given time.

___ 51. Automated systems should be used at the crews' discretion.

___ 52. Flying highly automated aircraft alters the way crewmembers transfer information.

___ 53. I try to use automation as much as possible during flight operations.

___ 54. It is difficult to know what FMC operations the other crewmember is performing.

### V. Flight Management Attitudes

The following items deal with attitudes regarding general flightdeck management. Please answer by writing beside each item a letter from the scale below.

| A | B | C | D | E |
|---|---|---|---|---|
| Disagree Strongly | Disagree Slightly | Neutral | Agree Slightly | Agree Strongly |

___ 55. The captain should take physical control and fly the aircraft in emergency and non-standard situations.

___ 56. Captains should encourage crewmember's questions about the flight.

___ 57. Even when fatigued, I perform effectively during critical times in a flight.

___ 58. Pilots should be aware of and sensitive to the personal problems of other crewmembers.

___ 59. The organization's rules should not be broken - even when the employee thinks it is in the organization's best interests.

___ 60. I let other crewmembers know when my workload is becoming (or about to become) excessive.

___ 61. I like my job.

___ 62. My decision making ability is as good in emergencies as in routine flying conditions.

___ 63. Effective crew coordination requires crewmembers to take into account the personalities of other crewmembers.

___ 64. To resolve conflicts, crewmembers should openly discuss their differences with each other.

___ 65. Junior crewmembers should not question the captain's decisions.

___ 66. A truly professional crewmember can leave personal problems behind when flying.

___ 67. I am proud to work for this organization.

___ 68. I am more likely to make judgment errors in an emergency.

___ 69. Crewmembers should not question the decisions or actions of the captain except when they threaten the flight's safety.

___ 70. Successful flight deck management is primarily a function of the captain's flying proficiency.

___ 71. If I perceive a problem with the flight, I will speak up, regardless of who might be affected.

___ 72. I am ashamed when I make a mistake in front of my other crewmembers.

___ 73. Written procedures are necessary for all in-flight situations.

___ 74. I am less effective when stressed or fatigued.

___ 75. Working for this organization is like being part of a large family.

___ 76. My performance is not adversely affected by working with a less capable crewmember.

___ 77. There are no circumstances (except total incapacitation) where the first officer should assume command of the aircraft.

___ 78. Personal problems can adversely affect my performance.

### VI. Work Values

Please answer the items below by writing beside each item a letter from the scale below.

| A | B | C | D | E |
|---|---|---|---|---|
| Of very little or no importance | Of little importance | Of moderate importance | Very important | Of utmost importance |

Please think of your *ideal* job - disregarding your present job. In choosing an *ideal* job, how important would it be to you to**:**

___ 79. Observe strict time limits for work projects?

___ 80. Have challenging tasks to do, from which you get a personal sense of accomplishment?

___ 81. Know everything about the job, to have no surprises?

___ 82. Have a changing work routine with new, unfamiliar tasks?

___ 83. Find the truth, the correct answer, the one solution?

*This completes the questionnaire. Thank you for your time.*

## APPENDIX B: UNADJUSTED MEANS FOR MANCOVAS AND ANCOVAS

Table B1. Range of Unadjusted Mean Scores (overall SD) for Three Training Evaluation Items, for Seven Airlines, Across All Fleets and For Two Selected Fleets.

| | | Airline | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Item | Fleet | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Item 26 | All fleets | 3.24-3.61 (1.32) | --- | 2.83-3.53 (1.37) | 4.03-4.25 (1.10) | --- | 2.94-4.57 (1.30) | 3.25-3.80 (1.29) |
| | B73-300 | --- | 3.60 (1.31) | --- | --- | --- | 3.63 (1.27) | 3.25 (1.33) |
| | B75/767 | 3.24 (1.33) | --- | 3.53 (1.29) | 4.25 (1.10) | 3.63 (1.05) | 3.16 (1.39) | 3.80 (1.19) |
| Item 29 | All fleets | 3.09-3.15 (1.24) | -- | 3.15-4.25 (1.39) | 3.84-3.97 (1.18) | --- | 2.87-4.14 (1.32) | 2.73-3.48 (1.37) |
| | B73-300 | --- | 3.58 (1.18) | --- | --- | --- | 3.39 (1.28) | 2.73 (1.27) |
| | B75/767 | 3.15 (1.28) | --- | 3.16 (1.43) | 3.97 (1.22) | 3.41 (1.14) | 2.87 (1.53) | 3.48 (1.38) |
| Item 30 | All fleets | 3.44-3.63 (1.21) | -- | 3.66-3.85 (1.25) | 2.88-3.08 (1.37) | --- | 2.90-3.87 (1.29) | 3.48-3.98 (1.25) |
| | B73-300 | --- | 3.50 (1.24) | --- | --- | --- | 3.69 (1.28) | 3.98 (1.10) |
| | B75/767 | 3.63 (1.14) | --- | 3.85 (1.19) | 3.04 (1.45) | 3.49 (1.20) | 3.87 (1.20) | 3.48 (1.35) |

*Note.* Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".
*Note.* Airlines without mean and SD ranges (Airlines 2 and 5) are represented in the sample by only one automated fleet.

Table B2. Unadjusted Mean Scores (SD) for Free-play and No Free-play Groups, Three Training Evaluation Items.

| Item | Free-play | No free-play |
|---|---|---|
| Item 26 | 3.83 (1.20) | 3.32 (1.46) |
| Item 29 | 3.59 (1.25) | 3.03 (1.43) |
| Item 30 | 3.36 (1.28) | 3.77 (1.32) |

*Note.* Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".

Table B3. Unadjusted Mean Scores (SD) Across Free-play Usefulness, Three Training Evaluation Items.

| Item | Not at all useful | Slightly useful | Somewhat useful | Very useful |
|---|---|---|---|---|
| Item 26 | 2.82 (1.08) | 3.33 (1.42) | 3.70 (1.29) | 3.99 (1.14) |
| Item 29 | 2.46 (0.93) | 3.17 (1.41) | 3.54 (1.19) | 3.78 (1.19) |
| Item 30 | 3.91(1.38) | 3.75 (1.29) | 3.47 (1.28) | 3.26 (1.28) |

*Note*. Item 26="My transition training adequately prepared me for flying this aircraft on the line". Item 29 = "I gained an adequate understanding of the FMS during transition training". Item 30 = "I learned most of what I need to know about this aircraft when I started flying it on the line".

Table B4. Unadjusted Means (SD) for Two Equipment Evaluation Scales, Seven Aircraft.

| Aircraft | Automation-Crew Synergy | Troubleshooting & Problem Solving |
|---|---|---|
| A320 | 35.08<br>(7.80) | 9.75<br>(1.29) |
| B737-300 | 33.92<br>(7.70) | 9.24<br>(1.49) |
| B747-400 | 35.83<br>(8.61) | 9.76<br>(1.46) |
| B757/767 | 35.41<br>(7.40) | 9.68<br>(1.43) |
| B777 | 37.31<br>(7.43) | 9.83<br>(1.42) |
| MD11 | 29.90<br>(9.12) | 9.65<br>(1.41) |
| MD80 | 31.29<br>(6.34) | 9.10<br>(1.36) |

Table B5. Means (SD) for Six Equipment Evaluation Items, Seven Aircraft Types.

| Aircraft | Item 14 | Item 11 | Item 9 | Item 24 | Item 25 | Item 17 |
|---|---|---|---|---|---|---|
| A320 | 2.62 (1.24) | 3.25 (1.34) | 1.85 (0.97) | 3.72 (1.09) | 3.72 (1.02) | 1.95 (1.24) |
| B737-300 | 2.79 (1.18) | 3.57 (1.24) | 1.82 (.91) | 3.72 (1.01) | 3.27 (1.03) | 1.63 (0.95) |
| B747-400 | 2.74 (1.22) | 3.59 (1.29) | 1.93 (1.13) | 3.76 (1.12) | 3.60 (1.17) | 1.84 (1.13) |
| B757/767 | 2.71 (1.17) | 3.54 (1.20) | 1.78 (.86) | 3.80 (0.98) | 3.64 (1.02) | 1.74 (1.06) |
| B777 | 2.77 (1.32) | 3.66 (1.24) | 1.75 (0.88) | 3.97 (0.90) | 3.73 (.97) | 1.73 (1.17) |
| MD11 | 2.65 (1.09) | 3.02 (1.30) | 2.75 (1.16) | 2.73 (1.25) | 3.22 (1.20) | 2.43 (1.33) |
| MD80 | 3.00 (1.07) | 3.32 (1.24) | 2.56 (1.13) | 3.17 (1.03) | 3.08 (0.95) | 1.75 (1.05) |

*Note.* Item 14 = “I get insufficient feedback regarding mode changes from the automated systems”. Item 11 = “I am always aware of changes to thrust level that are commanded by the autothrottle system”. Item 9 = “It is very difficult to get the information I need from the FMS in this aircraft”. Item 24 = “Information I may need from the FMS is logically presented”. Item 25 = “The error messages I receive from the FMS are useful in correcting a problem”. Item 17 = “It is difficult to transition from automated to manual flight in this aircraft”.

## REFERENCES


Air Line Pilots Association (1996, September). *Automated cockpits and pilot expectations: A guide for manufacturers and operators.* Air Line Pilots Association, Herndon, VA: Author.

Air Transport Association (1989, April). *National plan to enhance aviation safety through human factors improvements*. Washington, DC: Author.

Air Transport Association (in press). *Towards a model training program for FMC-generation aircraft.* Manuscript in preparation.

American Airlines (1996, October). *Advanced Aircraft Maneuvering Program.* Dallas, TX: Author.

Ajzen, I. & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888-918.

Amalberti, R. (in press). Automation in Aviation. In J. Wise, D. Garland, & D. Hopkin (Eds.). *Human Factors in Aviation.* Mahwah, NJ: Erlbaum.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19,* 775-779.

Bent, J. (1996). *Training for new technology*. Cathay Pacific Airways. Hong Kong: Author.

Billings, C.E. (1997). *Aviation automation: The search for a human-centered approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Billings, C.E., & Reynards, W.D. (1984). Human factors in aircraft incidents: Results of a 7-year study. *Aviation, Space, and Environmental Medicine, 55,* 960-965.

Bowers, C., Deaton, J., Oser, R., Prince, C., & Kolb, M. (1993). The impact of automation on crew coordination and performance. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 573-577) Columbus, Ohio: Ohio State University.

Byrne, B.M. (1989). *A primer of LISREL.* New York: Springer-Verlag.

Byrnes, R.E., & Black, R. (1993). Developing and implementing CRM programs: The Delta experience. In E.L. Wiener, B.G. Kanki, & R.L. Helmreich (Eds.), *Cockpit resource management* (pp. 421-443). San Diego: Academic Press.

Chidester, T.R., & Foushee, H.C. (1989). Leader personality and crew effectiveness: A full-mission simulation experiment. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Fifth International Symposium on Aviation Psychology.* Columbus, OH: Ohio State University.

Cooper, G.E., White, M.D. & Lauber, J.K. (1979). *Resource management on the flight deck* (NASA Report No. CP2120). Moffett Field, CA: NASA-Ames Research Center.

Danaher, J.W. (1980). Human error in ATC systems operations. *Human Factors, 22,* 535-545.

Degani, A., Shafto, M., & Kirlik, A. (in press). Mode usage in automated cockpits: Some initial observations. *Proceedings of the International Federation for Automatic Control (IFAC).* Amsterdam: Elsevier.

Degani, A., & Wiener, E.L. (1994). *On the design of flightdeck procedures.* (NASA Contractor Report No. 177642). Moffett Field, CA: NASA-Ames Research Center.

Dornheim, M.A. (1995). Dramatic incidents highlight mode problems in cockpits. *Aviation Week and Space Technology, 142(5),* 57-59.

Dornheim, M.A. (1996a). America West undergoes first post-Valujet inspection. *Aviation Week and Space Technology, 145(9)*, 22-23.

Dornheim, M.A. (1996b). Airlines improve automation training. *Aviation Week and Space Technology, 145(10),* 128-131.

Drury, C.G. (1996). Automation in quality control and maintenance. In Parasuraman, R. & Mouloua, M. (Eds.), *Automation and Human Performance* (pp. 407-426). Mahwah, NJ: Lawrence Erlbaum Associates.

Edwards, E. (1972). Man and machine: Systems for safety. *Proceedings of the British Airline Pilots Association Technical Symposium* (pp. 21-36). London: British Airline Pilots Association.

Endsley, M.R., & Kiris, E.O. (1994). The out-of-loop performance problem: Impact of level of automation and situation awareness. In M. Mouloua & R. Parasuraman (Eds.), *Human Performance in automated systems: Current research and trends* (pp. 50-56). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Fazio, R.H. (1986). How do attitudes guide behavior? In R.M. Sorrentino & E.T. Higgins (Eds.) *Handbook of motivation and cognition* (pp. 205-243). New York: Guilford.

Fazio, R.H., & Zanna, M.P. (1981). Direct experience and attitude-behavior consistency. In L. Berkowitz (Ed.), *Advances in experimental social psychology: Vol. 14* (pp. 161-202). New York: Academic Press.

Federal Aviation Administration (1991). *Advanced qualification program.* (Advisory Circular 120-54). Washington, DC: Author.

Federal Aviation Administration (1996). *FAA Human Factors Team report on the interfaces between flightcrews and modern flight deck systems.* Washington, DC: Author.

Folkerts, H.H., & Jorna, PG.A.M. (1994). *Pilot performance in automated cockpits: A comparison of moving and non-moving thrust levers.* National Aerospace Laboratory, The Netherlands: Author.

Foushee, H.C. & Manos, K.L. (1981). *Within-cockpit communication patterns and flight crew performance* (NASA Technical Paper 1875) Moffett Field, CA: NASA-Ames Research Center.

Forslin, J., Soderlund, J., & Zackrisson, B. (1979). Automation and work organization: A Swedish experience. In Forslin, J., Sarapata, A., & Whitehill, A.M. (Eds.), *Automation and industrial workers: A fifteen nation study: Vol. 1(1)* (pp. 31-64). Oxford: Pergamon Press.

Funk, K., Lyall, B., & Riley, V. (1995). Perceived human factors problems of flightdeck automation: A comparative analysis of flightdecks with varying levels of automation. (Phase 1 Final Report) Corvallis, OR: Oregon State University.

Gregorich, S.E., Helmreich, R.L., & Wilhelm, J.A. (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75,* 682-690.

Hansman, R.J., Midkiff, A. Vakil, S., Vaneck, T., Corwin, W., Irving, S., Irving, J., & Polson, P. (1995, January). *Mode awareness in advanced autoflight systems.* Paper presented to the FAA Human Factors Advisory Committee, Langley, VA.

Helmreich, R.L. (1984). Cockpit management attitudes. *Human Factors, 26,* 583-589.

Helmreich, R.L. (1987). Flight crew behaviour. *Social Behaviour, 2,* 63-72.

Helmreich, R.L., Chidester, T.R., Foushee, H.C., Gregorich, S., & Wilhelm, J.A. (1990). How effective is cockpit resource management training? *Flight Safety Digest, 9(5),* 1-17.

Helmreich, R.L., Foushee, H.C., Benson, R., & Russini, W. (1986). Cockpit resource management: Exploring the attitude-performance linkage. *Aviation, Space, and Environmental Medicine, 57,* 1198-1200.

Helmreich, R.L. & Foushee, H.C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E.L. Wiener, B.G. Kanki, & R.L. Helmreich (Eds.), *Cockpit resource management* (pp. 3-45). San Diego: Academic Press.

Helmreich, R.L., Hines, W.E., & Wilhelm, J.A. (1996). *Crew performance in advanced technology aircraft: Observations in four airlines.* (University of Texas Aerospace Crew Research Project Technical Report 96-8).

Helmreich, R.L., Merritt, A.C., & Sherman, P.J. (1996). Human factors and national culture. *ICAO Journal, 51(8),* 14-16.

Helmreich, R.L., Merritt, A.C., Sherman, P.J., Gregorich, S.E., & Wiener, E.L. (1993). *NASA/UT/FAA Flight management attitudes questionnaire (FMAQ): A new questionnaire containing items from the original CMAQ, cross-cultural items to tap national differences, and an attitudes toward automation scale.* (NASA/University of Texas/FAA Technical Report 93-4). Austin, TX: University of Texas.

Helmreich, R.L. & Sherman, P.J. (1994, October). Flightcrew perspectives on automation: A cross-cultural perspective. *Proceedings of the 7th ICAO Flight Safety and Human Factors Regional Seminar* (pp. 442-453). Addis Ababa, Ethiopia.

Helmreich, R.L., Wilhelm, J.A., & Gregorich, S.E. (1988). *Revised version of the cockpit management attitudes questionnaire (CMAQ) and CRM seminar evaluation form* (NASA/UT Technical Report 88-3). Austin, TX: University of Texas.

Hilburn, B. Molloy, R. Wong, D., & Parasuraman, R. (1993). Operator versus computer control of adaptive automation. In R.S. Jensen and D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 161-166). Columbus, Ohio: Ohio State University.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Beverly Hills, CA: Sage.

Hopkins, R. (1993, March 31-April 6). Backing up of approaches [Letter to the editor]. *Flight International,* p. 40.

Hulin, C.L. (1989). The role of communications in group operations of automated systems. *Proceedings of the Human Factors Society 33rd Annual Meeting,* 775-777.

Joreskog, K.G., & Sorbom, D. (1989). *LISREL 7: A guide to the program and applications.* Chicago: SPSS Inc.

Kanki, B. G., & Palmer, M.T. (1993). Communication and Crew Resource Management. In E.L. Wiener, B.G. Kanki, & R.L. Helmreich (Eds.), *Cockpit Resource Management*. San Diego, CA: Academic Press.

Kantowitz, B.H., & Campbell, J.L. (1996). Pilot workload and Flightdeck automation. In Parasuraman, R. & Mouloua, M. (Eds.), *Automation and Human Performance* (pp. 117-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Kantowitz, B.H., & Casper, P.A. (1988). Human workload in aviation. In Wiener, E.L., & Nagel, D.C. (Eds.), *Human Factors in Aviation*. San Diego: Academic Press.

Kalton, G. (1983). Introduction to survey sampling. *Sage University Paper series on Quantitative Applications in the Social Sciences (07-035),* Beverly Hills and London: Sage Publications.

Laming, J. (1993, June 9-15). Who controls the aircraft? [Letter to the editor*]*. *Flight International,* p. 140.

Lauber, J.K. (1984). Resource management in the cockpit. *Air Line Pilot, 53,* 20-33.

Lee, J.D., and Sanquist, T.F. (1996). Maritime automation. In Parasuraman, R. & Mouloua, M. (Eds.), *Automation and Human Performance* (pp. 365-384). Mahwah, NJ: Lawrence Erlbaum Associates.

Marsh, H.W., & Hocever, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin, 97,* 562-582.

Maurino, D. (1993). Cross-cultural perspectives in human factors training: The lessons from the ICAO human factors programme. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 606-611). Columbus, Ohio: Ohio State University.

McClumpha, A.J., James, M., Green, R.G., & Belyavin, A.J. (1991). Pilots' attitudes to cockpit automation. *Proceedings of the Human Factors Society 35th Annual Meeting,* 107-111.

Merritt, A.C. (1996). National culture and work attitudes in commercial aviation: A cross-cultural investigation. Unpublished doctoral dissertation. The University of Texas at Austin.

Merritt, A.C., & Helmreich, R.L. (1997). Manuscript in preparation.

Merritt, A.C., & Helmreich, R.L. (1996a). CRM: I hate it, what is it? (Error, stress, and culture). Paper presented at the Orient Airlines Association Air Safety Seminar, Jakarta, Indonesia, April 23-25, 1996.

Merritt A.C., & Helmreich, R.L. (1996b). Human factors on the flightdeck: The influence of national culture. *Journal of Cross-Cultural Psychology, 27(1),* 5-24.

Merritt, A.C., & Helmreich, R.L. (1995). Culture in the cockpit: A multi-airline study of pilot attitudes and values. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 676-681). Columbus, OH: Ohio State University.

Merritt, A.C., Helmreich, R.L., Wilhelm, J.A., & Sherman, P.J. (1996). *Flight Management Attitudes Questionnaire 2.0 (International) and 2.1 (USA/Anglo).* (Technical Report 96-04). Austin, TX: University of Texas.

Meshkati, N. (1996). Organizational and safety factors in automated oil and gas pipeline systems. In Parasuraman, R. & Mouloua, M. (Eds.), *Automation and Human Performance* (pp. 427-448). Mahwah, NJ: Lawrence Erlbaum Associates.

Mosier, K.L., Skitka, L.J., & Heers, S.T. (1995). Automation and accountability for performance. In R.S. Jensen and D. Neumeister (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 221-226). Columbus, Ohio: Ohio State University.

Muir, B.M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37(11),* 1905-1922.

Murphy, M.R. (1980). Analysis of eighty-four commercial aviation incidents: Implications for a resource management approach to crew training. *1980 Proceedings of the Annual Reliability and Maintainability Symposium,* 298-306.

National Academy of Sciences Panel on Human Factors in Air Traffic Control Automation. (in press). *Human operators and automation in air traffic control.* Washington, DC: National Academy Press.

National Transportation Safety Board. (1973). *Eastern Air Lines, Inc., L-1011, N310EA, Miami, Florida, December 29, 1972* (NTSB-AAR-73-14). Washington, DC: Author.

National Transportation Safety Board. (1979). *Aircraft accident report: United Airlines, Inc., McDonnell-Douglas, DC-8-61, N8083U, Portland, Oregon, December 28, 1978* (NTSB-AAR-79-7). Washington, DC: Author.

National Transportation Safety Board. (1982). *Aircraft accident report: Air Florida, Inc., Boeing 737-222, N62AF, collision with 14th Street Bridge, near Washington National Airport, Washington, D.C., January 13, 1982* (NTSB-AAR-82-8). Washington, DC: Author.

National Transportation Safety Board. (1984). *Scandinavian Airlines System Flight 901, McDonnell-Douglas DC-10-30, John F. Kennedy Airport, Jamaica, New York, February 28, 1984.* (NTSB-AAR-84-15). Washington, DC: Author.

National Transportation Safety Board. (1986). *China Airlines Boeing 747-SP, N4522V, 300 nautical miles northwest of San Fransisco, California, February 19, 1985.* (NTSB-AAR-86-3). Washington, DC: Author.

National Transportation Safety Board. (1994). *Aircraft accident report: Stall and loss of control on final approach, Atlantic Coast Airlines, Inc./United Express Flight 6291, Jetstream 4101, N304UE, Columbus, Ohio, January 7, 1994.* (NTSB-AAR-94-07). Washington, DC: Author.

National Transportation Safety Board. (1997). *Grounding of the Panamanian passenger ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts, June 10, 1995.* (NTSB-MAR-97-01). Washington, DC: Author.

Parasuraman, R. Mouloua, M. Molloy, R. & Hilburn, B. (1993). Adaptive function allocation reduces performance costs of static automation. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 178-181) Columbus, Ohio: Ohio State University.

Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation-induced “complacency”. *The International Journal of Aviation Psychology, 3(1),* 1-23.

Parasuraman, R., & Riley, V. (in press). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*.

Pew, R.W. (1969). Comments on ‘promotion of man’: Challenges in socio-technical systems; designs for the individual operator. In E.O Attinger (Ed.), *Global systems dynamics*. New York: Karger.

Ratwatte, S. (1997). Personal communication.

Riley, V. (1996). Operator reliance on automation: Theory and data. In Parasuraman, R. & Mouloua, M. (Eds.), *Automation and Human Performance* (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum Associates.

Rudisill, M. (1994). Flight crew experience with automation technologies on commercial transport flight decks. In M. Moloua & Parasuraman, R. (Eds.), *Human performance in automated systems: Current research and trends: Proceedings of the First Automation Technology and Human Performance Conference* (pp. 203-211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ruffell Smith, H.P. (1979). *A simulator study of the interaction of pilot workload with errors, vigilance, and decisions* (NASA Technical Memorandum 78482) Moffett Field, CA: NASA-Ames Research Center.

Sarter, N.B. (1996). Human errors are symptoms of a mismatch between pilots, machines, and the operating environment. *ICAO Journal, 51(8),* 9-10.

Sarter, N. B., & Woods, D.D. (1992). Pilot interaction with cockpit automation I: Operational experiences with the flight management system. *The International Journal of Aviation Psychology, 2(4),* 303-321.

Sarter, N. B., & Woods, D.D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots’ model and awareness of the flight management system. *The International Journal of Aviation Psychology, 4(1),* 1-28.

Sarter, N. B., & Woods, D.D. (1995). *Strong, silent and out of the loop: Properties of advanced cockpit automation and their impact on human-automation interaction.* (Cognitive Systems Engineering Laboratory Report CSEL 95-TR-01). Columbus, OH: The Ohio State University.

Sears, R.L. (1989). *Prevention strategies for the crew-caused accident.* Paper presented at the Egyptian Aircraft Accident Investigation and Prevention Seminar, Cairo, Egypt.

Sheridan, T.B. & Johannsen, G. (1976). *Monitoring behavior and supervisory control.* New York: Plenum Press.

Sherman, P.J., & Helmreich, R.L. (in press). Training for use of automation: The value of ‘free-play’ as a complement to traditional transition training. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Sherman, P.J., & Helmreich, R.L. (1995). Attitudes toward automation: The effect of national culture. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 682-687). Columbus, OH: Ohio State University.

Sherman, P.J., Helmreich, R.L., & Hines, W.E. (1997, March). The risks of automation: Some lessons from aviation, and implications for training and design. *Proceedings of the Workshop on Human Error and Systems Development, University of Glasgow (pp. 1-7).* Glasgow, Scotland: University of Glasgow.

Sherman, P.J., Helmreich, R.L. & Merritt, A.C. (in press). National culture and attitudes toward automation. *The International Journal of Aviation Psychology.*

Sherman, P.J., & Wiener, E.L. (1995). At the intersection of automation and culture. *Earth Space Review, 4(1),* 11-13.

Veinott, E.S. & Irwin, C.M. (1993). Analysis of communication in the standard vs. automated aircraft. In R.S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 584-588). Columbus, OH: Ohio State University.

Weener, E.F. (1991). *Key elements of accident avoidance.* Seattle: Boeing Commercial Airplane Company. Seattle, WA: Author.

Wickens, C.D. (1992). *Engineering psychology and human performance (2nd ed.).* New York: Harper Collins.

Wickens, C.D. (1994, September). *Designing for situation awareness and trust in automation.* Paper presented to the IFAC Conference on Integrated Systems Engineering, Baden-Baden, Germany.

Wiener, E.L. (1988). *Cockpit automation.* In E.L. Wiener & D.C. Nagel (Eds.), Human Factors in Aviation. San Diego: Academic Press.

Wiener, E.L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft.* (NASA Contractor Report 177528). NASA-Ames Research Center, Moffett Field, CA.

Wiener, E.L. (1993a). *Intervention strategies for the management of human error.* (NASA Contractor Report 4547). NASA-Ames Research Center, Moffett Field, CA.

Wiener, E.L. (1993b). Crew coordination and training in the advanced technology cockpit. In E.L. Wiener, B.G. Kanki, & R.L. Helmreich (Eds.), *Cockpit resource management* (pp. 199-229). San Diego: Academic Press. (1989).

Wiener, E.L., Chidester, T.R., Kanki, B.G., Palmer, E.A., Curry, R.E., & Gregorich, S.E. (1991). *The impact of cockpit automation on crew coordination and communication: Overview, LOFT evaluations, error severity, and questionnaire data* (NASA Contractor Report 177587). NASA-Ames Research Center, Moffett Field, CA.

Wiener E.L., & Curry, R.E. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23,* 995-1011.

Woods, D.D. (1997). Personal communication.

Woods, D.D. (1996). Decomposing automation: Apparent simplicity, real complexity. In Parasuraman, R. & Mouloua, M. (Eds.), *Automation and Human Performance* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Associates.